

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
 Organization  
 International Bureau



(43) International Publication Date  
 12 February 2004 (12.02.2004)

PCT

(10) International Publication Number  
**WO 2004/013327 A1**

- (51) International Patent Classification<sup>7</sup>: C12N 15/10, (74) Agents: BECKER, Philippe et al.; Cabinet Becker et Associés, 35, rue des Mathurins, F-75008 Paris (FR).  
 C07K 14/00, C07H 21/00
- (21) International Application Number: PCT/EP2003/007765
- (22) International Filing Date: 17 July 2003 (17.07.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
 02291871.8 24 July 2002 (24.07.2002) EP
- (71) Applicant (for all designated States except US): LIBRAGEN [FR/FR]; 11, Boulevard Einstein, F-69100 Villeurbanne (FR).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): NALIN, Renaud [FR/FR]; 54, Avenue du Général De Gaulle, F-69300 Caluire (FR). PUJIC, Petar [HR/FR]; 4, rue Chantemerle, F-69360 Serezin Du Rhone (FR). TUPHILE, Karine [FR/FR]; 181, rue de l'Oratoire, F-69300 Caluire (FR). GILLET, Benjamin [FR/FR]; 46, rue de la Thibaudière, F-69007 Lyon (FR).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
 — with international search report  
 — before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR THE EXPRESSION OF UNKNOWN ENVIRONMENTAL DNA INTO ADAPTED HOST CELLS

(57) Abstract: The present invention relates to methods for the identification or cloning of polynucleotides encoding a selected phenotype, particularly from environmental DNA. In a specific embodiment, the method comprises (i) cloning environmental DNA fragments into Ecoli cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

BEST AVAILABLE COPY

WO 2004/013327 A1

## Method for the expression of unknown environmental DNA into adapted host cells

### Introduction and Background

5 The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic,  
10 agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural  
15 products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which  
20 are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc...) have not been cultivated because  
25 their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

30 New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation

processes through their long evolution. These approaches are called "Metagenomic" because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

- 5 Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic has been firstly developed for the identification of new bacterial phylum (Pace .1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics
- 10 relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve a first step of gene selection (generally using PCR) before cloning. In the case of protein production, the cloning vector used are preferentially also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene
- 15 in a given bacterial host strain.
- More recent developments of metagenomic consider the total metagenome cloned without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds
- 20 of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been cloned (more than 30 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways for the identification of new metabolites resulting from multi enzymatic synthesis. This last point is of
- 25 particular interest for bacterial metagenomic libraries since, most the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria. Nevertheless, the heterologous expression of a whole biosynthetic pathways (large DNA fragment) needs a much more improved system than a simple expression vector to have a full and stable expression.

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154 and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries have been established in *E. coli* which is the most efficient cloning system. However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as Actinomyces, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from Actinomicetes and some specific substrates for proteins activity are not present in *E. coli*.

The stable maintenance of large foreign DNA fragments (> 10Kb) into a selected host cell is one of the key points for academic research or applied industrial purposes. Usually, the vector carrying the foreign DNA is maintained by cultivating the host cells in a medium with a vector-specific selective pressure (resistance to an antibiotic for example). However, when large foreign DNA fragments are cloned and/or expressed, their propagation and/or expression require energy, which is not allocated for cell growth anymore. As a consequence of this new resource allocation (nutrients/energy), it is not unusual to have a genetic rearrangement of the foreign DNA (deletion, modification etc...) as a recombinant cell reaction. This results in the modification of the foreign genetic information and in the loss of DNA functionality. This can be observed without

any loss of the selective pressure carried by the vector. As a result, the recombinant clone is no more exploitable for genetic or functional analysis.

Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow stable maintenance and propagation of large foreign DNAs into host cells for production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the environmental DNAs.

#### Summary of the Invention

The present invention discloses improved tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify and produce novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow genetic and functional analyses of particular selected clones in any appropriate system, thereby making possible the stable maintenance and propagation, the analysis and/or the expression of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E. coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying

said selected cloning vectors to allow a transfer of said vectors into a selected host cell and integration of the polynucleotide contained in said vectors into the genome of the selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

5

An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target  
10 polynucleotide construct allowing (i) transfer of the environmental DNA into a selected host cell distinct from *E. coli*, (ii) integration of the environmental DNA into the genome of a selected host cell, and (iii) stable maintenance and propagation of the environmental DNA into the selected host cell.

15 A further object of this invention is a method of producing modified libraries of polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

20 The modification in the library or cloning vector is typically obtained by targeted insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

The integration into the genome of the selected host cell is typically obtained by site  
25 specific integration or by homologous or heterologous DNA/DNA recombination.

The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify new genes and isolate new metabolites, drugs, enzymes, antibiotics, etc.

30

### Legend to the Figures

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene conferring resistance to apramycin (Apra) fig.1b.

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains an origin of transfer (oriT) and a gene conferring resistance to apramycin (Apra). Direction of DNA transfer at oriT is shown by an arrow.

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains  $\Phi$ C31 integrase gene and attachment DNA sequence for site specific integration, origin of transfer and gene for selection. The orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable construct into selective gene marker carried on original vector is shown on fig. 4b. In this event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Complete annotated DNA sequences of fosmid clones FS3-124 (a, SEQ ID NO: 1) and FS3-135 (b, SEQ ID NO: 2).

Figure 6 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

Figure 7: Schematic Map of pPSB vector (top) and transposable elements (bottom). The transposable element has 720 bp DNA of the *amyE* gene from *B. subtilis* in addition to *att-int-oriT-Apra*.

- 5 Figure 8: Plasmid pPSBery (top) and transposable element (bottom) carrying selective marker *ery* AM, for resistance to erythromycin, and part of *amy E* gene for homologous recombination in *B. subtilis*.

- 10 Figure 9: Integrase  $\Phi$  C31 was deleted from pPSBery plasmid. Resulting plasmid is pPSBery-DI (top). Transposable element contains *Apra* and *ery* AM genes for selection, *oriT* origin of transfer and a part of *amyE* gene for integration in to *amyE* locus of *B. subtilis* chromosome (bottom).

- 15 Figure 10: Map of pTn5-7 AOI plasmid. Transposable element has ends of *tn5* (ME) and *tn7* (T7 R, T7 L) transposons.

### Detailed Description of the Invention

- 20 The invention provides novel strategies, methods and products for generating and analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a
- 25 transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

- 30 In a most preferred embodiment, the methods allow stable transfer and propagation of large environmental nucleic acids in a selected host following initial selection. Such



methods comprise (i) selecting cloning vectors in a library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and integration of the polynucleotide which they contain into a selected host cell genome, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

#### Library of polynucleotides

- 10 The term "library of polynucleotides" designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such unknown (or uncharacterized) polynucleotides, the library may further include known sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the polynucleotides are generally cloned into cloning vectors, allowing their maintenance and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.
- 25 The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed "metagenomic libraries" since they contain nucleic acids derived from whole genomes of mixed populations of microorganisms.
- 30 The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable)

microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g., industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral extracts, sand, natural excrements, meteorits etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

10

Environmental samples usually contain various species of (uncharacterized, uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms, eubacteria, archaeobacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles, psychrophiles, psychrotophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic); as well as nucleic acids and organic materials. The sample may also contain different animal cells: mammalian cells; insect cells, etc (arising from larvae, feces, etc.).

25

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension,

30

paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

- 5 The term "environmental DNA" designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al.* (Chemistry & Biology 5(10), 1998, R245), Rondon *et al.* (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller  
10 et al (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al.* (Applied and Environm. Microbiol., 65, 1999, 5409).

- In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic  
15 acids, such as environmental RNAs, for instance.

- The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar  
20 size, to produce homogenous libraries.

### Cloning Vectors

- As indicated, the polynucleotides are contained or cloned into cloning vectors. These  
25 vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors  
30 essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified

after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into fosmids, BAC and P1 derivatives.

- 5 As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will  
10 be adapted later during the process.

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically contain, in addition to the polynucleotide insert, an origin of replication functional in a  
15 selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or maintenance in other cell types.

20

In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC and p15A, for instance. Many *E. coli* cloning vectors are commercially available and/or  
25 can be constructed using available regulatory sequences.

#### Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide  
30 library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various

techniques, such as molecular screening, protein expression, functional screening, etc. A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as hybridisation, amplification, sequencing, etc. Preferably, molecular screening comprises the selection of clones in a library which contain, in their sequence, a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

10 In a first variant, the selection is made by contacting the cloning vectors in the library with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

15

In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library which lead to a positive amplification product are then selected.

20

In this regard, the present application provides new primers designed in conserved motives of the  $\beta$ -keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative polyketide synthase (PKS) genes or domains. These primers have the following degenerated sequence :

25

Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)

Antisense primer : 5'- GCBBSRYTCDATSGGRTCSCC -3' (SEQ ID NO: 4)

wherein :

R is A or G

30 S is G or C

Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

5 A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

10 Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to study their function and/or produce novel compounds or metabolites.

The invention now enables such further analysis and uses, by allowing a modification of  
15 the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

20 Modification of the cloning vectors

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or characterisation of cloned DNA fragments, the invention now allows to modify  
25 specifically the cloning vectors to transfer, integrate into the genome, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or heterologous host cells, and include, but are not limited to, for example *Streptomyces*, *Nocardia*, *Bacillus*, fungi, yeasts, etc.

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By "specific" is meant that the modification occurs at a pre-determined location in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

10

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring the selected property(ies) to the cloning vector.

15 The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, integrate into the genome, maintain, express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in combination(s).

20

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

25 Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an origin of transfer or "ori T". The presence of such an oriT in a vector allows said vector to be transferred into a desired host cell.

30

In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

The structure of various oriT has been reported in the art (Guiney *et al.* 1983; Zechner *et al.* 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K( $\alpha$ ).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence (SEQ ID NO: 5) :

```
gatctGTGATGTACTTCACCAGCTCCGCGAAGTCGCTCTTCTTGATTGGAGCGCATGGG
GACGTGCTTGGAATCACGCGCACCCCCGCGCGTTTATAGCGGCTAAAAAAGTCAT
GGCTCTGCCCTCGGGCGGACCACGCCCATCATGACCTTGCCAAGCTCGTCC TGCTTC
15 TCTTCGATCTTCGCCAGCAGGGCGAGGATCGTGGCATCACCGAACC GCGCCGTGCG
CGGGTCGTGCGGTGAGCCAGAGTTTCAGCAGGCGCCAGGCGGCCAGGT CGCCAT
TGATGCGGGCCAGCTCGCGGACGTGCTCATAGTCCACGACGCCC GTGATT TTGTAGC
CCTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCGCCGCCTTT
TCCTCAATCGCTCTTCGTTTCGTCTGGAAGGCAGTACACCTTGATAGGTGGGCTGCCC
20 TTCCTGGTTGGCTTGTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGG
TAGCCGGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGTGCGAATAA
GGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCC
CGGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGGCAAATC
CTGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCG
25 AAGCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT
```

The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause conjugative transfer. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

In an other particular variant, the target polynucleotide construct comprises a genetic element allowing integration of the vector (or of the polynucleotide contained therein) into the genome of the selected host cell.



A donor DNA is permanently or stably maintained and expressed in a selected recipient cell if it is integrated into the recipient cell's genome or if it contains elements that allow autonomous replication in said cell. In a most preferred embodiment, the vector is modified to allow transfer and integration of the polynucleotide into the host cell genome.

Integration is a preferred way of ensuring stable expression. Integration can be obtained by physical recombination. Recombination can be homologous, e.g., between two homologous DNA sequences, or illegitimate, where recombination occurs between two non-homologous DNAs. As a particular example, integration of donor DNA into the chromosome of the recipient can be mediated by host recombination repair system or by site-specific recombination. Another well-studied process that can transfer and integrate genes is transduction by bacterial viruses, such as  $\lambda$  and  $\phi$ C31. In a phage-infected bacterial cell, fragments of the host DNA are occasionally packaged into phage particles and can then be transferred to a recipient cell. Integration into the recipient cell's genome is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid encoding an integrase functional in the selected host cell. More preferably, the integrase is selected from  $\lambda$  and  $\phi$ C31 integrases. In a specific embodiment, the polynucleotide construct comprises a nucleic acid encoding an integrase having or comprising all or a functional part of the following sequence of the  $\phi$ C31 integrase (SEQ ID NO: 6) :

```
AGATCTCCCGTACTGACGGACACACCGAAGCCCCGGCGGCAACCCTCAGCGGATGC
25 CCCGGGGCTTCACGTTTTCCAGGTCAGAAGCGGTTTTTCGGGAGTAGTGCCCCAACT
GGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGACATGACACAAG
GGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAG
CGCGAGAATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGGCGGTTTCAGGTTCTGTC
30 GGGCATTTCAGCGAAGCGCCGGGCACGTTCGGCGTTCGGGACGGCGGAGCGCCCGGA
GTTTCAACGCATCCTGAACGAATGCCGCGCCGGGCGGCTCAACATGATCATTGTCT
ATGACGTGTCGCGCTTCTCGCGCCTGAAGGTCATGGACGCGATTCCGATTGTCTCGG
AATTGCTCGCCCTGGGCGTGACGATTGTTTCCACTCAGGAAGGCGTCTTCCGGCAGG
GAAACGTCATGGACCTGATTACCTGATTATGCGGCTCGACGCGTCGCACAAAGAA
35 TCTTCGCTGAAGTCGGCGAAGATTCTCGACACGAAGAACCTTCAGCGCGAATTGGG
CGGGTACGTTCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTTCGGAGACGAAGG
```

AGATCACGCGCAACGGCCGAATGGTCAATGTCGTCATCAACAAGCTTGCGCACTCG  
ACCACTCCCCTTACCGGACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGGTGG  
CGTGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGCAGTCAAGCCGCCAT  
TCACCCGGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGA  
5 CCCGGGGCGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACC  
GTTATGCGAATCCTTCGGGACCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTAC  
AAGAAGAAGCCGGACGGCACGCCGACCACGAAGATTGAGGGTTACCGCATTCAGCG  
CGACCCGATCACGCTCCGGCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCG  
CTGAGTGGTATGAGCTTCAGGCGTGGTTGGACGGCAGGGGGCGCGGCAAGGGGGCTT  
10 TCCCGGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTGGCGC  
CGTCATGACTTCGAAGCGCGGGGAAGAATCGATCAAGGACTCTTACCGCTGCCGTC  
GCCGGAAGGTGGTCGACCCGTCCGCACCTGGGCAGCACGAAGGCACGTGCAACGTC  
AGCATGGCGGCACTCGACAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAGGCA  
CGCCGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCCGACGCTTCG  
15 GCAAGCTCACTGAGGCGCTGAGAAGAGCGGCGAACGGGCGAACCTTGTTGCGGAG  
CGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGGCAGGCGC  
GTACGACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGCAGCGCTGACGC  
TCCGGCAGCAAGGGGGCGGAAGAGCGGCTTGCCGAACCTTGAAGCCGCCGAAGCCCCG  
AAGCTTCCCCTTGACCAATGGTTCCCCGAAGACGCCGACGCTGACCCGACCGGCCCT  
20 AAGTCGTGGTGGGGGCGCGCGTCAGTAGACGACAAGCGCGTGTTTCGTCGGGCTCTT  
CGTAGACAAGATCGTTGTCACGAAGTCGACTACGGGCAGGGGGCAGGGAACGCCCA  
TCGAGAAGCGCGCTTCGATCACGTGGGCGAAGCCGCCGACCGACGACGACGAAGAC  
GACGCCCAGGACGGCACGGAAGACGTAGCGGCGTAGCGAGACACCCG

25 The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause integration. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

30 In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

35

In an other variant, the target polynucleotide construct comprises an origin of replication specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAM $\beta$ 1, pSa, 2 $\mu$ m circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

40

In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person, depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac,  $\lambda$ PL, GAL, AOX, hsp-70, etc.

The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate selection means.

In a particular, preferred embodiment, the target polynucleotide construct is contained in or comprises a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host cell.

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the experimental section. The transposable nucleic acid construct may further comprise an integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated using inverted repeats obtained from transposons or transposable elements such as Tn5, Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from

transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- left arm of pPAOI6 transposon (SEQ ID NO: 7)

5 CTGTCTCTTATACACATCTCAACCATCATCGATGAATTTTCTCGGGTGTCTCGCATA  
TTGGCTCGAATTCGAGCTCGGTACCC

- right arm of transposon pPAOI6 (SEQ ID NO: 8)

10 GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCCAACGACTACGCACTAGCC  
AACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring resistance to antibiotics, such as apramycine, chloramphenicol, ampiciline, kanamycine,  
15 spectinomycine, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene and the modification step ii) comprises:

20 . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and  
. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

25

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

30 It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative units. Also, transposition can be carried out by *in vitro* process.

### Analysis of the polynucleotides

In step (iii) of the process, the polynucleotides may be analysed by various methods, including by genetic, biochemical, chemical or phenotypical approaches, which are well known per se in the art. Analysis occurs upon transfer and, optionally, expression of the polynucleotides into the selected host cell.

In this regard, the modified cloning vectors can be transferred into the selected host cell by a variety of techniques known in the art, including by transformation, electroporation, transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the target polynucleotide construct comprises an oriT and the modified vectors are transferred into the selected host cells by conjugative transfer. In this embodiment, the cloning vector and the selected host cells are co-cultivated and the recombinant host cells are selected and isolated.

The selected host cell may be any type of cell or microorganism, including, without limitation, *Streptomyces*, *E. coli*, *Salmonella*, *Bacillus*, *Yeast*, *fungi*, etc.

One of the objectives of the invention is to be able to analyse environmental DNAs of unknown cellular origin into different host expression systems. In order to analyse the potentiality of the DNAs at the transcription and/or translation levels and to have much more probabilities to have a DNA expression, it is important to have the possibility to test different host expression systems.

25

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce, for instance, an increase in doubling time, morphological modifications, pigments production, etc., which can be related either directly to the expression of foreign DNA or by combinatorial biology of the foreign DNA and the biology of the expression host systems. The new phenotypes can be analysed by all techniques known in the art such genetic, biochemical, chemical, phenotypic approaches etc.

30

In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to  
5 produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning  
10 vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These  
15 polynucleotides contain the sequence of PKS genes and other genes that encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in Figure 5. The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOs: 1 or 2), their  
20 complementary strand, or a functional variant thereof. A part of the above sequences includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an CDS). In this respect, SEQ ID NOs: 1 and 2 comprise several novel open reading frames encoding  
25 novel polypeptides involved in biosynthetic pathways. These coding sequences are identified in Figures 5a and 5b.

25

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides (CDS) 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ;  
2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ;  
10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505  
30 - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 -

22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 (figure 5a) or a complementary strand thereof.

In an other specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 (figure 5b) or a complementary strand thereof.

Variants of these sequences include any naturally-occurring variant comprising or several nucleotide substitutions ; sequences variants resulting from the degeneracy of the genetic code, as well as synthetic variants coding for functional polypeptides. Variants include any sequence that hybridise under high stringent conditions, as disclosed for instance in Sambrook et al., to any of the above sequences, and encode a functional polypeptide. The invention also include any nucleic acid molecule encoding a polypeptide comprising all or a fragment of an amino acid sequence encoded by a polynucleotide as disclosed above. Preferably, the fragment comprises at least 10 consecutive amino acid residues, more preferably at least 20, even more preferably at least 30.

The invention also relates to any vector comprising these polynucleotide sequences. These sequences may be DNA or RNA, preferably DNA, even more preferably double-stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above. The invention also relates to a method of producing such polypeptides by recombinant techniques, comprising expressing a polynucleotide as defined above in any suitable host cell and recovering the encoded polypeptide. The invention also relates to a recombinant host cell comprising a polynucleotide or a vector as defined above.



An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into the genome of a selected host cell distinct from *E. coli*.

The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of mutants by mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

Further aspects and advantages of the invention will be disclosed in the following examples, which should be regarded as illustrative and not limiting the scope of this application.

### Experimental Section

#### 25 A - From *E.coli* to *Streptomyces*

In this work, we constructed a fosmid library in *E. coli* from total DNA prepared directly from soil. The library has been screened for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene identification, inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramicyne. Conjugative derivatives were constructed by incorporating  
5 origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from  $\phi$  C31 phage, including attP attachment site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed  
10 transposon was tested for inactivation of the genes cloned into fosmids. Obtained mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

### Materials and methods

#### 15 Bacterial strains, plasmids and growth conditions.

*E. coli* DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification.

Unless specifically described, all DNA manipulations were performed according to  
20 Sambrook, J., *et al*, (1989).

#### Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

#### Fosmid DNA extraction and Purification

25

Fosmids DNA containing soil library were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing 12,5  $\mu\text{g ml}^{-1}$  of chloramphenicol. Cultures were grown at

37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

## 5 Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic acids sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the  $\beta$ -keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the  $\beta$ -keto acyl synthase gene. Lib1F (sense  
10 primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSRYTCDATSGGRTCSCC -3') were deduced from  $\beta$ -keto acyl synthase peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR  
15 systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis*, *Streptomyces lividans*, *Streptomyces ambofaciens* and *Ralstonia solanocearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

20

## PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO,  $MgCl_2$ , Primers and DNA template quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50 $\mu$ l) contained 250 $\mu$ M of dNTP, 5  
25 mM  $MgCl_2$  final, 2,5% DMSO, 1X PCR buffer, 0,75  $\mu$ M of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50 $\mu$ l) contained 250 $\mu$ M of dNTP, 5 mM  $MgCl_2$  final, 5% DMSO, 1X PCR buffer, 0,75  $\mu$ M of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives  
30 clones in 96 microtiter plates, 25  $\mu$ l of each bacteria culture were used as template and PCR conditions were the same as above. Thermocycling program was : a denaturation

step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives  
5 clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes. The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

#### PCR products analysis

10 PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman  
15 Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

#### Sequencing of the identified fosmid insert DNA and sequence analysis

Fosmids inserts were sequenced using either a transposon-mediated and by shotgun  
20 subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel electrophoresis for fragments ranging from 1 to 3  
25 Kbs and cloned into Bluescript vector according to Sambrook *et al.* (1989).

Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

### Sequencing

Sequencing reactions were performed with 1 µg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions proposed by supplier. Ten µL of reaction products were precipitated using 4 µL of solution containing 1.5 M NaOAc, 50 mM EDTA and 60 µL cold 95% ethanol/dH<sub>2</sub>O from -20 °C. The pellet was washed 2 times with 200 µL 70% ethanol/dH<sub>2</sub>O, vacuum dried and dissolved in 40 µL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

### 10 Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF ( d-CCCTAAGATCTGGTTCATGTGCAGCTCCATC, SEQ ID NO: 9) and AmR ( d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID NO: 10). One hundred microlitre reaction were performed containing 0.1µM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 µM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase (NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final extension was performed at 72 °C for 7 min.

PCR product was purified using GFX DNA purification kit (Amersham), then digested by *Bgl* II and *Sma* I restriction enzymes. A 941 bp *Bgl*II/*Sma*I fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were amplified via PCR using primers oriTF (d-GCGGTAGATCTGTGATGTACTTCACCAGCTCC, SEQ ID NO: 11) and oriTR

(TAGTACCCGGGGATCCGACGGATCTTTTCCGCTGCAT, SEQ ID NO: 12). PCR conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma*I restriction enzymes. *Bgl*II/*Sma*I DNA fragment was subjected to purification after gel electrophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested by *Bam* HI and *Sma* I restriction enzymes.

5  
10  
15  
 $\phi$  C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATCGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was purified by GFX kit (Amersham) and phosphorylated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorylated DNA fragment was cloned in to pPL1 vector opened with *Sma*I restriction enzyme (NEB) and dephosphorylated by calf alkaline phosphatase (NEB). DH10B *E. coli* was transformed with ligation mixture using Bio Rad Pulsing apparatus and protocols provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl*II restriction enzymes followed by digestion by Bean mung nuclease (NEB). Linearised plasmid was self ligated and transformed in to DH10B cells.

#### Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

#### Mutagenesis

25  
30  
Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the transposed mixture by electroporation into competent DH10B *E. coli* strain.

### Conjugation *E. coli*- *Streptomyces lividans* TK24

Conjugation experiments were done using  $6 \times 10^6$  *E. coli* S17.1 cells containing conjugative plasmids or fosmids. The *E. coli* cells were grown in LB media with adequate antibiotic. The cells were collected by centrifugation, washed two times using same volume of LB media and concentrated to  $10^8$  cells/ml and overlaid on LB plates containing  $2 \times 10^6$  pregeminated *Streptomyces lividans* TK24 spores. The cells mixture were grown overnight at  $30^\circ\text{C}$  and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlaid using top agar containing NAL (nalidixic acid) and the appropriate antibiotic. Plates were incubated for 4 days at  $30^\circ\text{C}$  and transformant streptomyces colonies were isolated on HT medium (Pridham *et al.* 1957) containing NAL and the same appropriate antibiotic.

## Results

### Construction of a Transposon Tn <Apra>.

*E. coli* aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre). Advantage of this selective marker allows positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a, 1b.

### Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from plasmid RP4 into pP1 vector producing pP1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced into specific *E. coli* S17.1 strain that carry RP4 plasmid integrated into chromosome. In conjugation experiment between donor strain S17.1 carrying pP1 plasmid and DH10 *E. coli* receptor we obtained DH10B strain carrying pP1 plasmid. This data shows that cloned oriT fragment is functional in the pP1 plasmid. Plasmid pP1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned

could be then transferred by conjugation in to another host. Another advantage of this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal  
5 gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage  $\phi$ C31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli* S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for  
10 plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene for resistance to apramycine. Conjugative transfer of pPAOI6 gene in to *S. lividans* strain is confirmed by resistance to apramycine or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the  $\phi$ C31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

15 pPAOI6 transposon was cloned into the *EcoRV* site of plasmid pGPS3 (New England Biolabs). This construction allows transposition not only by transposase Tn5 but also using Transposase ABC (New England Biolabs). Resulting plasmid pTn5-7AOI is shown on Figure 10.

The goal of these constructions was to produce transposons that is further used in  
20 functional analysis of the metagenomic DNA library from the soil that were constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA. The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents  
25 of the library was screened by using a PCR approach for the presence of the genes involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive clones indicate that the DNA library contains very little sequence redundancy limited to



one and that the sequences were found to be new and very diverse in comparison to gene databases (data not shown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and 69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites (Fig 5 a and b). No specific phenotype was observed for the two clones in *E. coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycin as selective antibiotic. Obtained mutants were then tested on LB plates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into *Streptomyces lividans* TK24 strain. Fosmid DNA was prepared from mutants and transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of the conjugative fosmid into *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in comparison to the control (Figure 6).

#### **B - From *E. coli* to *Bacillus subtilis***

Plasmid pPSB was constructed as follows : A part of amyE gene from *B. subtilis* was amplified by PCR using primers amyE-BamHI : atcgaggatcctgaggactctgaacccg (SEQ ID NO: 15) and amyE-EcoRI : cgactgaattcagatctagcgtgaaattccgtctgc (SEQ ID NO: 16). DNA fragment was digested by EcoRI and BamHI restriction enzymes and ligated into EcoRI, BglII site of pPAOI6 plasmid. Transposable element contains Tn<amyE-int  $\phi$ C31-oriT-apra> is shown on figure 7 with plasmid pPSB.

Plasmid pPSBery (Figure 8) was constructed by cloning erm AM gene into pPSB plasmid. A 1140 bp *Sau3A* DNA fragment containing ery AM gene with his own promoter was cloned from plasmid pMUTIN (Wagner *et al.* 1998) into the Bam HI site of plasmid pPSB. Orientation of ery gene was confirmed by sequencing. Transposable  
5 element contains Tn < amyE-int  $\phi$ C31-ery-oriT-apra> (Figure 8).

Plasmid pPSBery- $\Delta$ I was obtained after *Sma* I digestion and self-ligation of core plasmid (Figure 9). In this construction,  $\phi$ C31 integrase gene was deleted from transposable element. New transposon is Tn<amyE-ery-oriT-apra> (Figure 9). All transposons could  
10 be released as linear DNA by *Pvu*II digestion from plasmids mentioned above and transposed by transposase Tn 5 (Epicentre) in vitro.

The selection of transposed elements was done using 100  $\mu$ g/ml erythromycine or 40 $\mu$ g/ml apramycine in *E coli* or 0.3  $\mu$ g/ml erythromycine in *B subtilis*. DNA was  
15 transformed by electro-transformation into electrocompetent *E coli* strains or by competence into *B. subtilis*. Integration of imported DNA into amy E locus of *B. subtilis* chromosome was confirmed using pPSBery- $\Delta$ I plasmid. Integration was confirmed by PCR using plasmid-specific and amyE locus-specific primers. Fifteen eryR *B. subtilis* clones were tested by PCR. All transformants showed integration at amyE locus of *B*  
20 *subtilis* chromosome, confirming the functionality of the method and constructs.

References

- Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible  
5 involvement of an IS 140 element in its expression. Mol Gen Genet. 193 (1), 179-187.
- Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. Proc Natl Acad Sci USA, 80: 3595-8.
- 10 T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296, 1127-1129.
- Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between *E. coli* and *Streptomyces* species. J. Bacteriol. 171, 3583-3585.
- 15 Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276: 734-740.
- Pridham, T.G., Anderson P., Foley, C., Lindenfesler, L.A., Hesseltine, C.W., and Benetdict R.C. (1957). A selection of media for maintenance and taxonomic study of  
20 *streptomyces*. Antibiotics Annual 1956-1957, 947-953.
- Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by intergeneric conjugation. FEMS Microbiol Lett. 162(1):47-52.
- 25 Sambrook, J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2<sup>nd</sup> edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.
- Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. Appl.  
30 Environ Microbiol. 56, 782-787.

- Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. *J. Industr. Microbiol.* 17, 170-178.
- Torsvik, V., Daae, F.L, Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* 64. 53-62.
- Wagner, V., Dervyn, E. and Ehrlich S. D. 1998. A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology.* 144, 3097 – 3104.
- 10 Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In *The need for New antibiotics: Possible Ways Forward. Fifty –Third Symposium of the Society for General Microbiology*, Bath. England, UK, ( Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.
- 15 Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) *The horizontal gene pool*. Harwood Academic Publishers.

CLAIMS

1. A method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer and integration of said vectors and/or of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.
2. The method of claim 1, wherein the library comprises a plurality of unknown polynucleotides.
3. The method of claim 1 or 2, wherein the library comprises a plurality of environmental DNA fragments.
4. The method of any one of the preceding claims, wherein the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, fosmids, P1 or BAC.
5. The method of any one of the preceding claims, wherein the selected vectors are modified by targeted insertion, into the vector, of a target polynucleotide construct.
6. The method of claim 5, wherein the targeted insertion is performed in a region of the vector distinct from the polynucleotide.
7. The method of claim 5 or 6, wherein the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.
8. The method of claim 7, wherein the origin of transfer is functional in *E. coli* host cells.

9. The method of claim 8, wherein the origin of transfer is selected from RP4, pTiC58, F, RSF1010, ColE1 and R6K( $\alpha$ ).
10. The method of any one of claims 5 to 9, wherein the target polynucleotide construct  
5 comprises an integrase functional in the selected host cell.
11. The method of claim 10, wherein the integrase is selected from  $\phi$ C31.
12. The method of any one of claims 6 to 11, wherein the target polynucleotide construct  
10 comprises a transcriptional promoter functional in the selected host cell.
13. The method of any one of claims 5 to 12, wherein the target polynucleotide construct comprises a transposable nucleic acid construct.
14. The method of claim 13, wherein the transposable nucleic acid comprises, flanked by  
15 two inverted repeats, the target polynucleotide construct and a marker gene.
15. The method of any one of the preceding claims, wherein the cloning vector comprises a first marker gene and wherein, in step ii), the cloning vector is modified by:  
20 . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and  
. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.  
25
16. The method of any one of the preceding claims, wherein, in step (i), the cloning vectors which contain a polynucleotide having a particular characteristic are selected by molecular screening.
17. The method of any one of the preceding claims, wherein, in step (iii), the modified  
30 cloning vectors are transferred into the selected host cell by conjugative transfer.

18. The method of any one of the preceding claims, wherein, in step (iii), polynucleotides are analysed by determining the phenotype or properties of the host cell upon transfer or expression of the modified vector.

5

19. A method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

15

20. A transposable nucleic acid construct, wherein said construct comprises an origin of transfer and elements for integration and selection in a selected host cell genome flanked by two inverted repeats.

21. A library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into a selected host cell distinct from *E. coli*.

25

22. A polynucleotide sequence comprising all or part of SEQ ID NOs: 1 or 2, or of their complementary strand.

23. An oligonucleotide comprising SEQ ID NO: 3 or 4.

30

1/39

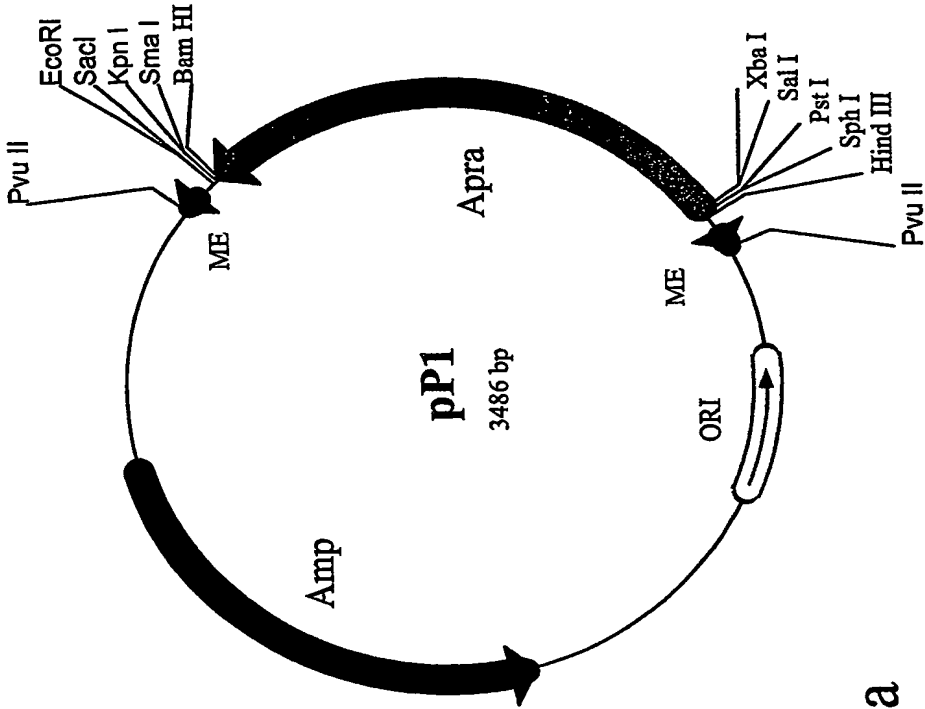


Figure 1a



2/39

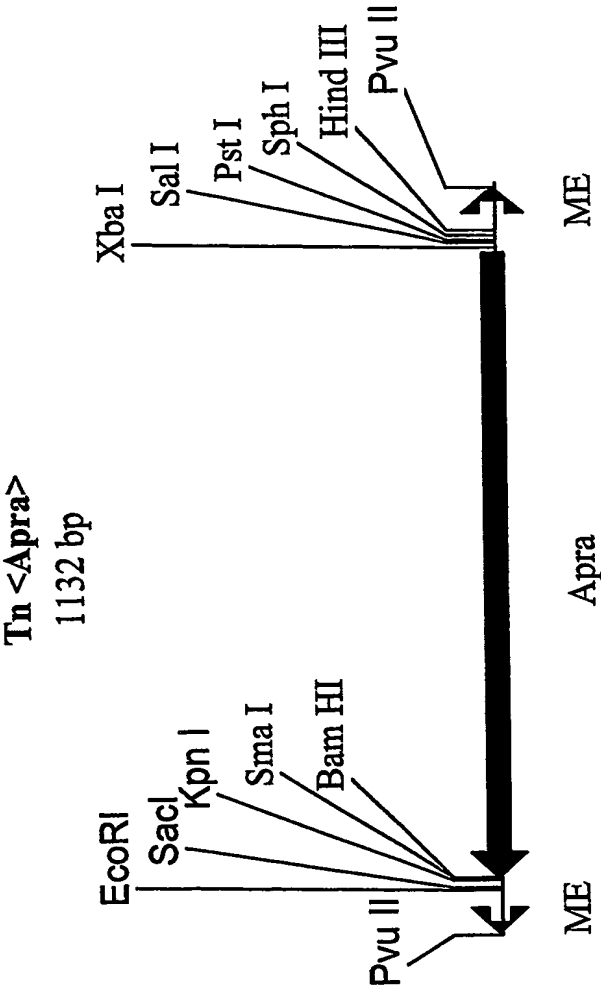


Figure 1b

3/39

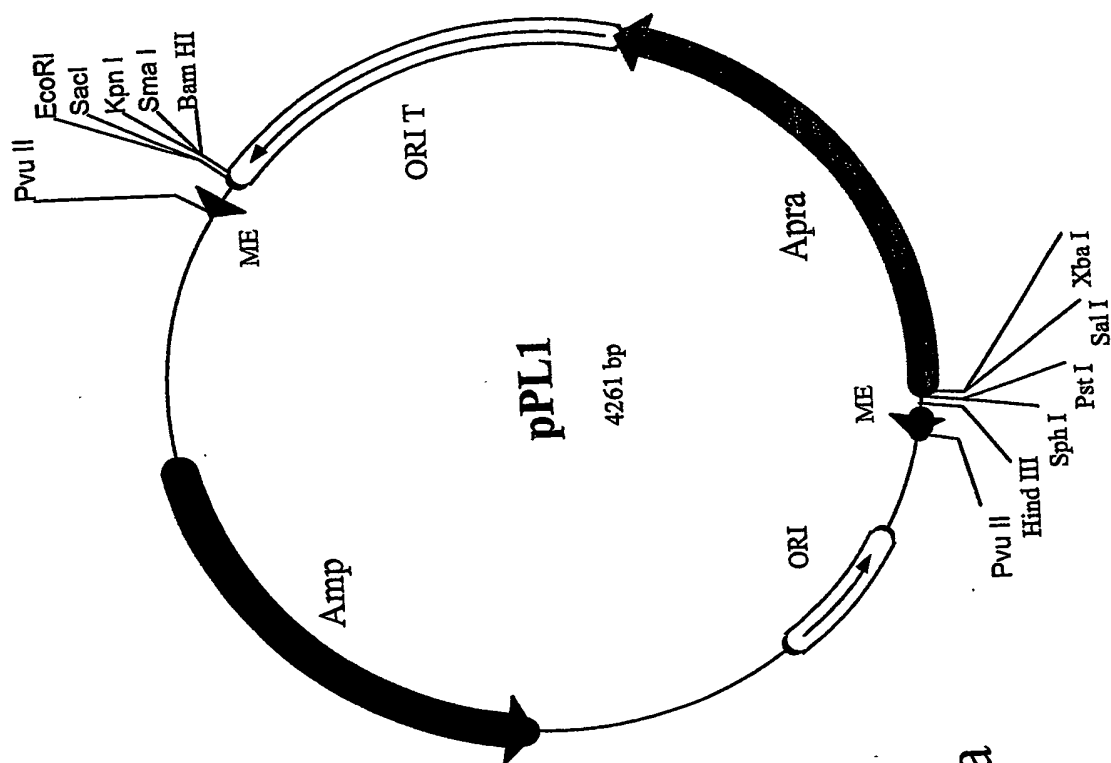


Figure 2a

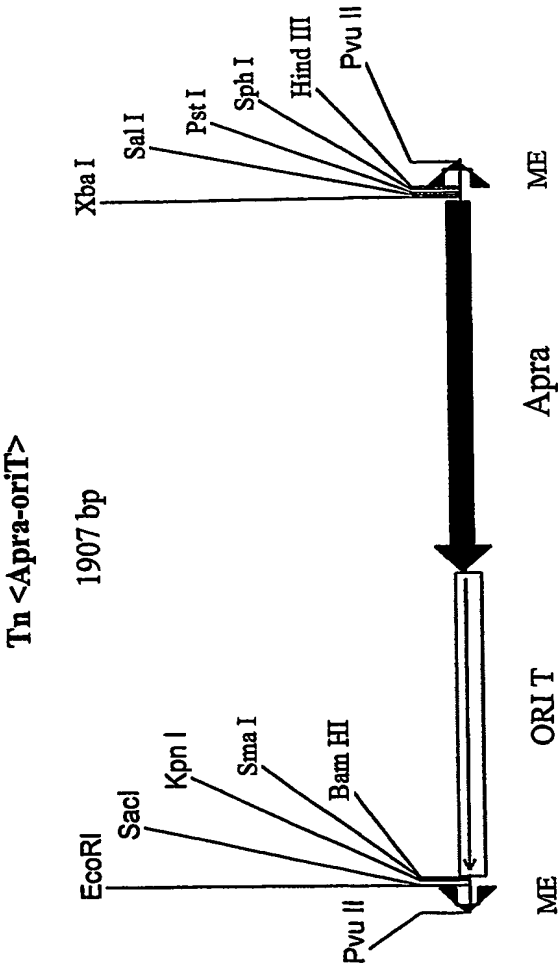


Figure 2b

5/39

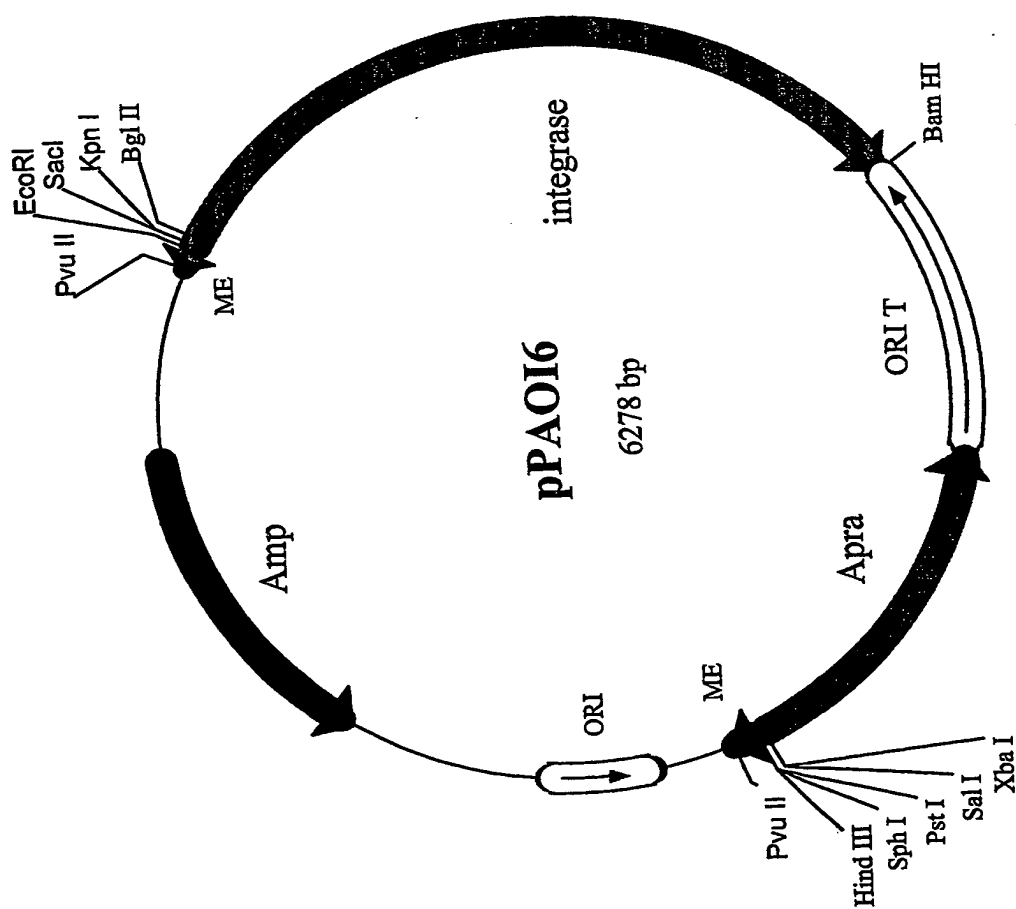


Figure 3a

6/39

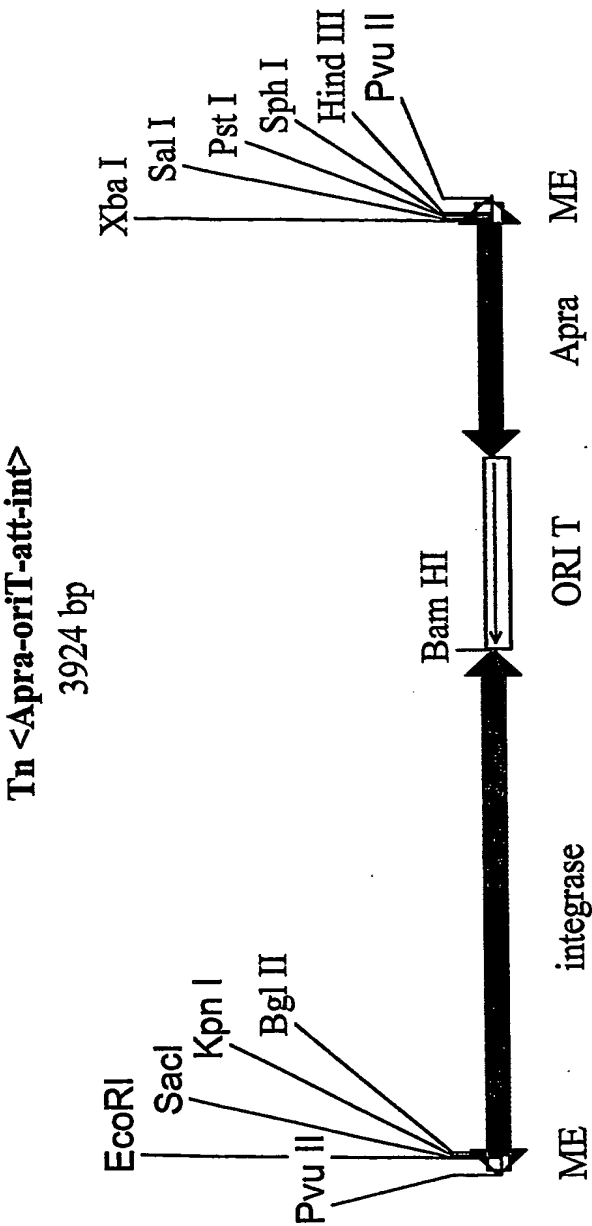


Figure 3b

7/39

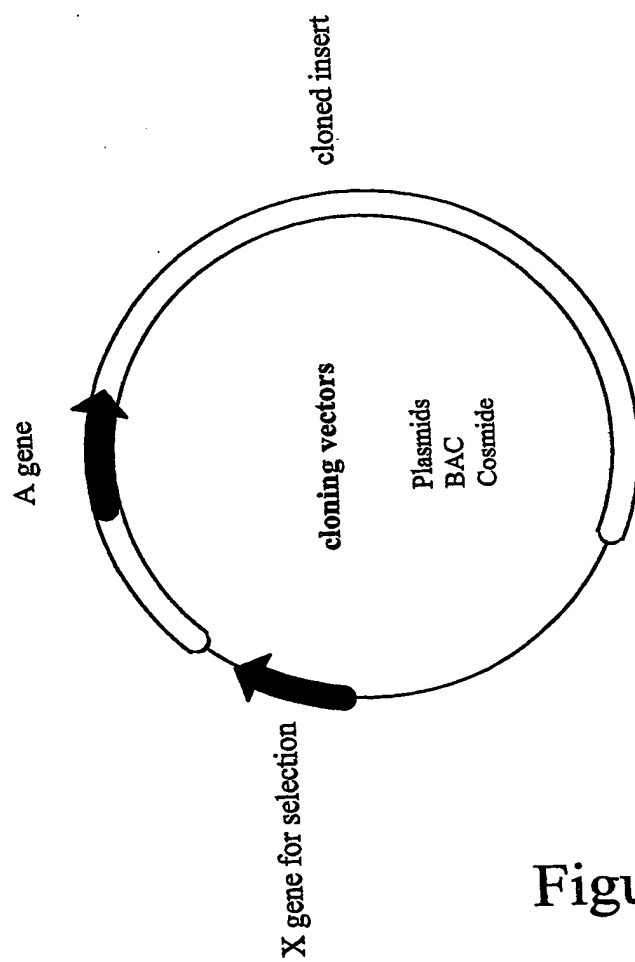
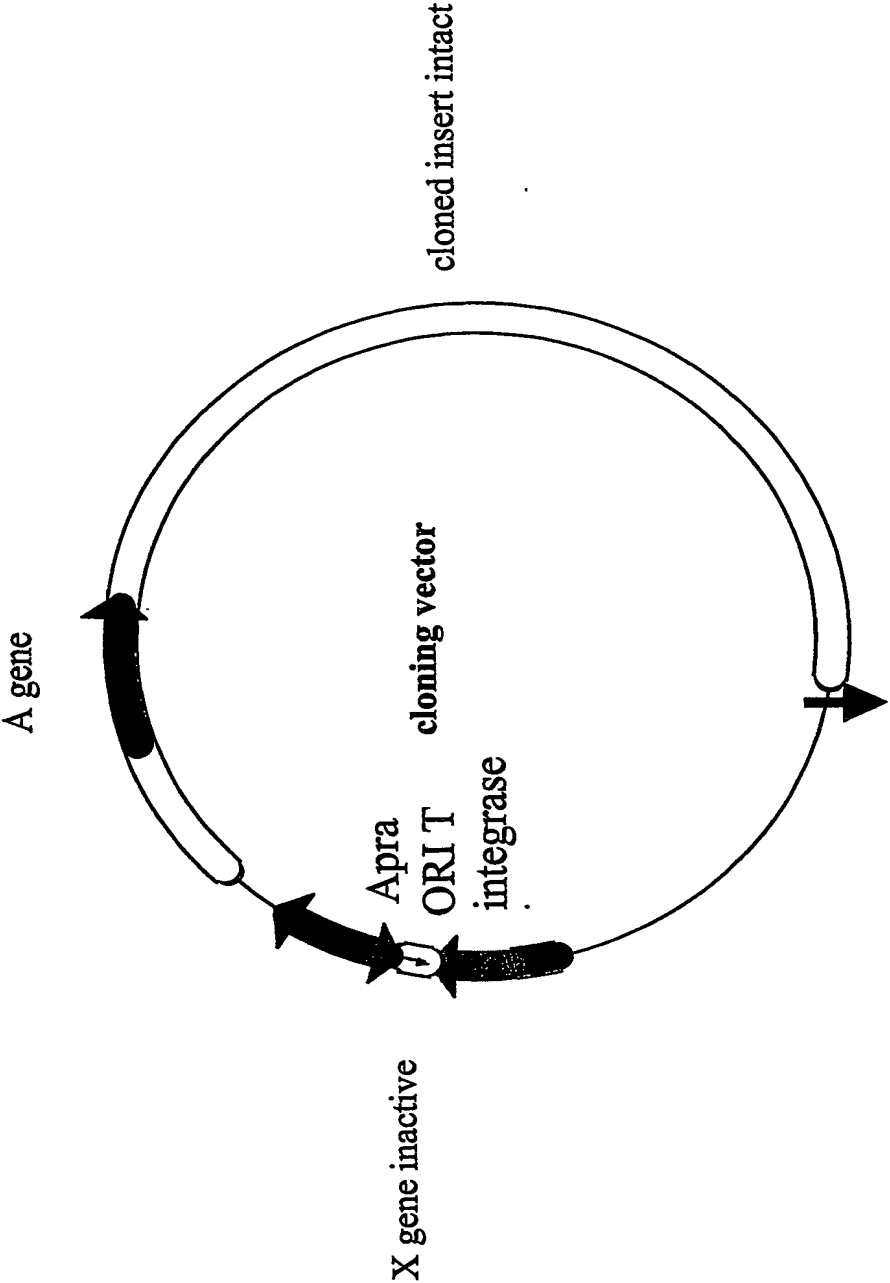


Figure 4a

8/39



Conjugative and integrative plasmid

Figure 4b

9/39

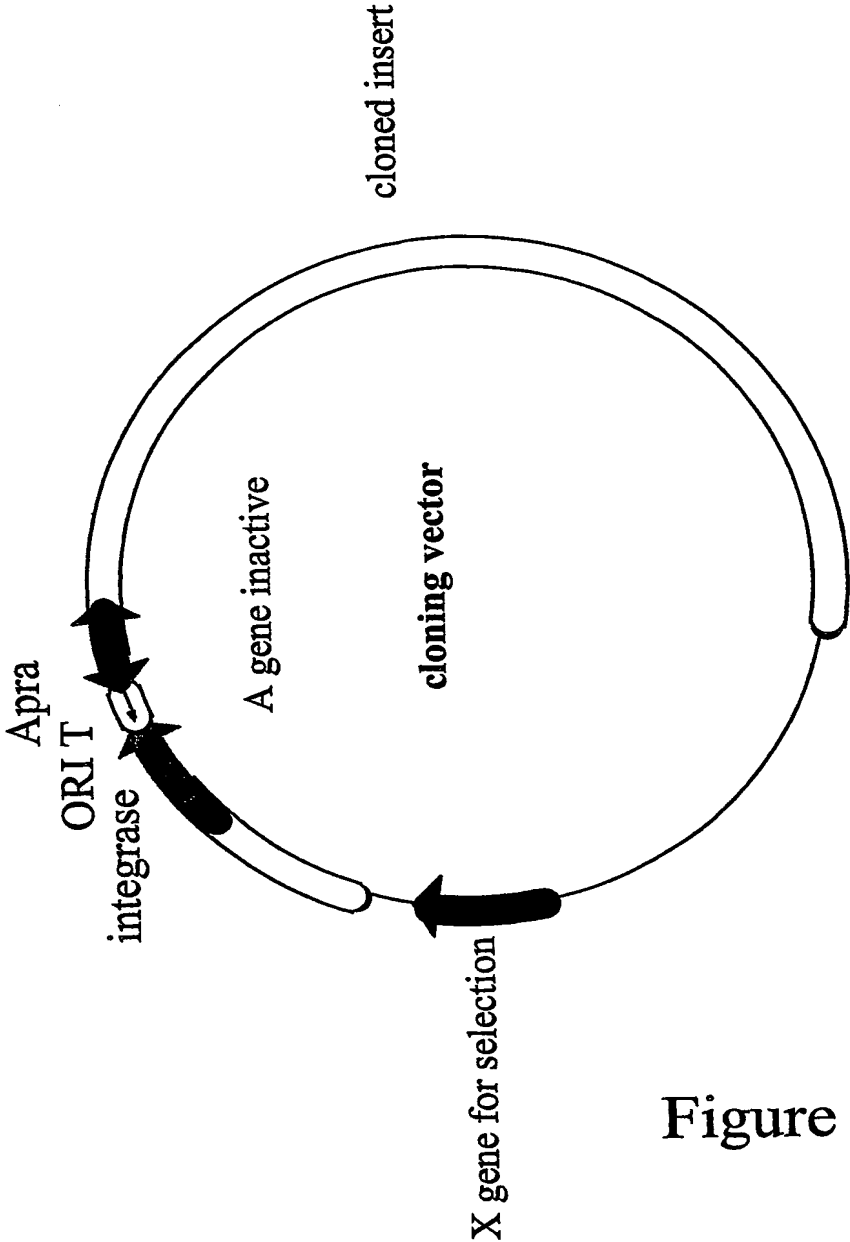


Figure 4c



## Figure 5a

```

CDS      76..1134
         /note="ABC_transportr"
         /gene="TAP2 PROTEIN"
         /blastp_match="Oryzias latipes"
         /blast_score= 0.002
CDS      complement(1096..2430)
         /note="none"
         /blastp_match="Anabaena sp"
         /gene="ALR1117 PROTEIN"
         /blast_score=2e-18
CDS      1178..1624
         /note="Gram_pos_anchor"
         /gene="CELL WALL SURFACE ANCHOR"
         /blast_score=1e-04
         /blastp_match="Streptococcus pneumoniae"
CDS      complement(2506..3567)
         /note="CONSERVED"
         /gene="HYPOTHETICAL PROTEIN"
         /blast_score=0.019
         /blastp_match="Deinococcus radiodurans"
CDS      complement(2906..4222)
         /note="glycosyl transferase"
         /gene="lipopolysaccharide"
         /blast_score=2e-23
CDS      complement(4092..5321)
         /note="glycosyl transferase"
         /gene="glycosyl transferase"
         /blast_score=1e-15
CDS      complement(6337..8502)
         /note="PUTATIVE"
         /gene="GLUTAMINE AMIDOTRANSFERASE"
         /blastp_match=" Bordetella bronchiseptica"
         /blast_score=1e-16
CDS      complement(8181..9530)
         /note="none"
         /gene="MEMBRANE PROTEIN"
         /blast_score=0.035
CDS      complement(9531..10721)
         /note="NOEC Transmembrane"
         /gene="NODULATION PROTEIN"
         /blastp_match="Azorhizobium caulinodans"
         /blast_score=3e-07
CDS      complement(10504..11274)
         /note="PUTATIVE"
         /gene="HYDROLASE"
         /blastp_match="Streptomyces coelicolor"
         /blast_score=4e-14
CDS      12874..13689
         /note="HYPOTHETICAL Meth-transf"
         /gene="PROTEIN PA1088"
         /blast_score=2e-06
         /blastp_match="Pseudomonas aeruginosa"
CDS      14195..15976
         /note="PUTATIVE, Glyco_transf"
         /note=""
         /gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS"
         /blast_score=2e-06
         /blastp_match="Vibrio cholerae"
CDS      15427..16512
         /note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE
BIOSYNTHESIS"
         /gene="PHOSPHOHEPTOSE ISOMERASE "
         /blast_score=3e-17
         /blastp_match="Helicobacter pylori"
CDS      15579..16253

```

/note="none"  
/gene="PHOSPHOHEPTOSE ISOMERASE"  
/blast\_score=2e-22  
/blastp\_match="Neisseria meningitidis"  
CDS complement(16505..17656)  
/note="BIOSYNTHESIS PUTATIVE"  
/gene="LIPOPOLYSACCHARIDE"  
/blast\_score=2e-17  
/blastp\_match="Thermotoga maritima"  
/pfam\_match="Glycos\_transf\_1"  
CDS complement(17657..18697)  
/note="none"  
/gene="ALR3073 PROTEIN"  
/blast\_score=6e-27  
/blastp\_match="Anabaena sp"  
CDS complement(18615..19304)  
/note="none"  
/gene="ALR4487 PROTEIN"  
/blast\_score=8e-07  
/blastp\_match="Anabaena sp"  
CDS complement(19301..20596)  
/note="ATP\_GTP\_A"  
/gene="ABC TRANSPORTER"  
/blast\_score=3e-61  
/blastp\_match="Synechocystis sp"  
CDS complement(20535..21476)  
/note="PERMEASE COMPONENT"  
/gene="POLYSACCHARIDE ABC TRANSPORTER"  
/blast\_score=6e-41  
/blastp\_match="Clostridium acetobutylicum"  
CDS complement(22025..22951)  
/note="involved in the synthesis of a polysaccharide capsule ?"  
/gene="32.3 KDA PROTEIN"  
/blast\_score=3e-17  
/blastp\_match="Sphingomonas sp"  
CDS 23155..26523  
/note="peptide syntase"  
/gene="mcyA, mcyB and mcyC"  
/blastp\_match="Microcystis aeruginosa"  
/blast\_score=0.0  
CDS 26409..34433  
/note="polyketide syntase et peptide syntase"  
/gene="mcyD, mcyE, mcyF and mcyG"  
/blastp\_match="Microcystis aeruginosa"  
CDS 34418..37500  
/note="CYSTATIN"  
/gene="PEPTIDE SYNTHETASE"  
/blast\_score=0.0  
/blastp\_match="Anabaena sp"  
CDS 35359..37500  
/note="gene cluster"  
/gene="nostopeptolide biosynthetic"  
/blast\_score=2e-41  
/blastp\_match="Nostoc sp"

Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;

gatcgggtgcg	ggcctcttcg	ctattacgcc	agctggcgaa	agggggatgt	gctgcaaggc	60
gattaagtgtg	ggtaacgcca	gggttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	120
aattgtaata	cgactcacta	tagggcgaa	tgcagctcgg	tacccgggga	tcccacgtac	180
cacgagctca	tctggaagag	cccgcctccg	acacccgagg	aatggaagcg	cggcgctcgtg	240
agaccgtaca	cgcagaagcg	tctcgtcgcc	ttcgttgact	ggcttgcgac	gacgatgaag	300
ctggatctca	caggatgtt	cgccgcagcg	agctcgatgg	gcggatcggg	cgcgatcatg	360
ctcgcgattc	gctatccgc	acgattcgcg	tggaccgtgt	cgtgggtcgg	cgtccacgtg	420
cccgccgact	ctccgcaatt	cacatcgctg	tacgagctgg	tgtacggccg	gcccgacttg	480
aaggtgccgt	tgcagaacgg	cacgcgggtc	tgggatcatt	ttagtgacgt	ctggtacctg	540
cggcagcatc	cggagcagga	catcgggttc	atcacgttct	cgaacggcaa	gaacgactcg	600
gcgatcggct	ggcgccaggc	cgtcgaattc	ctgaagacgc	tgcaggagac	caggcaaccg	660
cacctgttcg	tctggggcca	ggaggggacac	ggccagcgcg	cgaagatgcc	ggaaggtggc	720

ggcgaacggg	agatgccgct	cgacctcaga	acgaaccaga	gcctgccggc	gttcagccgt	780
tgacagctcg	acgacgatcc	gggagacggg	tcggattcga	gcggcgctcc	ggccggggcag	840
atcaactcct	acgtcacctg	gcagcctgac	tcgatcgtcg	acgaaccggg	ccgatggagc	900
gtcgtcatca	agctggcgaa	tcgaccgccc	cggtcgcccg	ccgcagtcga	cgtcacgccc	960
cgccggctgc	agcagttcaa	gctcaagccc	ggcgatcagg	tgacatggac	gaactccgcg	1020
ggcaacaccg	tcgtgcagcg	cgccgaagcg	gtcggcgacc	gggtgggtct	cgtcacgctc	1080
ccgcagatgc	aggtctcaac	gggtggaaac	aggatcctgg	tgagcaggaa	ctaggcgggc	1140
ggcgtcctcg	acgagccggc	gcatcgggcg	cagttgatgt	gcccacgaaa	gctggctgag	1200
gtccgcgcgc	gcggcgatgt	cccgctgcgc	ctgcggcggc	tttgccagca	aatccagcac	1260
agccgcccgg	aacgcctcga	ctgattccgc	cacggagcac	gccggccgtg	cgatgggtgg	1320
aagtccccct	gcgacggccg	gggtcacgac	gttcggggagc	cccgcggcaa	tcgcttcgag	1380
cactttgttc	tgacgcctc	tcgcgatcgc	aatcgggcgc	accgcgatcg	ccgatcgcca	1440
gagatactcg	cgtacctcgc	ggaccgcgc	ggtcacttca	atcgtgggat	cggccgcgag	1500
tgccccacc	gcggcggtcg	gggtccatgcc	cacgagcatc	agccgcgcgt	ccggggcggc	1560
gcgaaccacg	gacggccaga	cctctcgcgc	gaaccagagc	gccgcgcgct	cgttgggctc	1620
gtagccgaac	acgccgcaga	agacgacccg	cgttcggcg	gccgggtggtc	ctttcggcgc	1680
aaacatctcg	acgtcgatgc	cgttcggcac	gaccgacgcc	gtgaggccgc	ccgtgacttc	1740
ttcgagcagc	aggcgctcgc	gctcgtcac	gaccgtgggtc	gccgcgcgcg	gccgcatcgc	1800
gactttctcg	aagcggagca	agcggccggc	ctcgcggcgg	aacagccagc	ctttcggcca	1860
cgccgcccgtc	gacgccagcg	cgcgccattt	ctcggagtcg	acgtccacca	tgtccaggac	1920
gaaggggatc	ccgcccgagg	tcggctcgca	ggcgtagcgg	gccatcccgc	tgcatagcgc	1980
gagcaccgcg	tcggggcgcg	tcgcggcaat	ccgtcgttcg	agcacctggt	agatctccgg	2040
cgaatgcagc	aggacgtgcg	tgagggttcg	atcgcccgcg	agcgccagac	cgtcgcgat	2100
gagattccgg	ccgcggcgga	cgcgcacgac	gtcggtcgac	gccgtcacgc	cggcgagatc	2160
cgctcgatgc	gaccactcct	cgtcgtcgtg	cgcgagcgag	acgaggtgca	cgtcggcgga	2220
tcgtgcgagc	gtgtgaatga	gatgaaacgc	gcgaatgcga	tcgccccggt	tcggggcgta	2280
gggcagccgg	tgtgtgagca	gaaggactct	caccgtgcgg	catccgatcg	ctctcggatg	2340
acgcggggcg	gtacgcccgc	ggcgatcacg	cgggcgggaa	tcggcgcggt	cacgaccgac	2400
ccgcgcgcaa	tcacgggaatc	ccgaccgacg	tcagccatga	ccacggcggc	gctgccgatc	2460
caggtgccgg	cgccgatccg	caccatccgc	ggtgtgcctg	gctgctcacg	gatcgggcgc	2520
gtgaggtcgc	tcgtgccgtg	cgtgtccggt	ccgctgggga	tgtggacgcc	ggcgccgacc	2580
aggacatcgc	gctccagggtg	gacccagccg	agatggcatc	cgggtcccac	gtacacatgc	2640
tcgtcagctc	tcgcgcgggt	ctgcgaaaag	atggtgcga	aggcgaccgt	gaccgacggg	2700
tcgcagtggt	caagcaacgc	gccgaggaaa	gccgcgcgaa	ggtattgtcc	gatgacgccg	2760
gggatgagcg	agagccactg	cgtcgaccct	tcgagggcc	ggtcggcgcc	gaggaagggc	2820
cgccgcacgg	agtaggagag	caaggcgggc	aggacgagca	ccagcgcat	gccacggggc	2880
agtcctctgc	cggcgctcct	gaggctcacg	cccactgcag	cgcgcgcccc	gcgtcggacc	2940
gggtccgcta	tcgtgtgacg	agcagtcgt	agagcgcttc	cactttgcgc	atccgcgct	3000
cgaacgataa	ctcttcttcg	atccggcgct	gcgcgcgaat	cgcacgggtg	cgtccggccg	3060
cgggattggc	catcgctcgc	tcgatcgcgt	ccgcgaggcc	tgacgcccgg	ccgggctgaa	3120
tgaccagggc	gtgcgtgccc	ttcgacacga	gctcggccgt	gccccgggce	gcggtcgcga	3180
cgatgggtgt	ctcgaaacgc	atggcttcga	gcacggcggt	gggcgtccct	tcgtagtccg	3240
acgactgcac	gagcaggtcc	agcccgtgat	ggacggccga	gacgtcggac	gtgtggccga	3300
ggaaacgcga	ttctccgggc	ccgagctcgc	ggcgccgctg	ttctcagagc	ggacgcccga	3360
ggctgccgtc	gccggcgatg	acgagacgga	ggcgccggca	cttcgcctga	agggcgccac	3420
aggcttcgat	gagcaagtcg	aatcgtttct	gcggctcgag	acggccaacc	gcgccgatga	3480
cgaagtgcgt	ggcctcgaat	cccagtcctt	cacgcgcgcg	ggcctcagat	cgtggatcgc	3540
gtcggaaaag	atggtggtcg	atgccattcg	ggatcgtgac	gaccgatcc	ggcgtccgc	3600
ctttggccag	gagctcccgc	cgaaatctgt	cggacacggc	aatgacggca	gggaagcggg	3660
cgagaatccg	gcgatcgagc	gggtagtaga	cgggtgcgtt	gcgaaacgaa	tggccggtcc	3720
atccatgggc	cgtggccagc	ggaatcacgc	gctcggcgcg	tcccaggagc	agggcgagga	3780
gatcgggtct	gtactcgtgc	gcgtgcacga	tgtcgatgcg	cagatctttc	acgagcgcg	3840
ggagcttcga	ccagattccg	gtgtcgaacg	aatgcttctc	ccggacctcg	acgtagtcca	3900
tgccggccgc	tttcgcccgt	tcgtccatgc	cgaacaccgg	atcgcgctgg	tcgcggatgt	3960
agcagacgg	gatcgcgtac	ttggagcgat	cggcctgcgc	cgttccgagc	agaatcgtct	4020
tctccggtcc	gccgcccgtg	ccccgcacgc	tcgggagctc	gagcacgcgc	acgggacgcg	4080
cgattcccgg	cctaggcgcg	gcgtctggca	gcatggcccc	cccggccgac	gagcgcgatg	4140
gccccgggca	tatcgtacgt	gccggccacg	acgagcgact	gtgcgatctc	gacgagtcct	4200
actgcgccgc	gtccggcccg	gatcgcgcg	gctgcgcgca	agccggcgaa	accggcaagc	4260
agcaggagcg	cggatccgac	aatccaccac	ccgcgcgcgt	gccacagcgc	gaggcccagc	4320
ggaatcgaga	cgagcgccag	cagatcgagg	accgggatga	tcacgctggg	cagatcccgc	4380
atggcgagcc	ggtgacggaa	gctcacgcgc	aggttgcgcg	ggccgcgcca	cagctcgccg	4440
aaaaacacgg	cgcgaagcga	tcggggatcg	ccctggtgca	cgtgcgggag	ccgcgggtcc	4500
gacaggatgc	ggccgcgcgc	ccggcgacgc	cgtggtcaga	tatccacgtc	ttcacaggtc	4560
tcgaggctgg	tatcgaagcc	accggcgcg	tcgaaggcct	cgcgcttcac	ggcaagattg	4620
ccgctcggga	gccagtcac	gtcttcgggtg	ctctgcccac	gacggcgaaa	cgagtcgtac	4680
gcgcgctgca	cccagttgcc	gtcacccggga	gcttcgtact	gcgcgcgggc	agcgacgacc	4740
cccggggctc	ggagcagatc	gacggcgacc	accggctcagc	gcgggtcgat	ctcatgatcg	4800
gcgtcaacga	acgcacgac	gtcgccctcc	gccgaccggg	ccgcgcgggt	gcgaagcgtc	4860

gccacgttca	cgtcggggcag	cgacagcact	cgcgcggtatc	cgcgctctgc	cacggcacgc	4920
gagtcgtcgg	tcgaccggtt	gtccgccacg	acgatctcgt	accagccagc	cgcgccctcg	4980
gtgcatgaa	tcgacgccag	gcaacggctg	aggtgcgcgc	cgccgttctt	gacggggatg	5040
acgaacgaca	cgcgcgccgc	cacggcaatt	ttcgcggtca	tcgatttctg	agcgcgcccg	5100
gctttcgtag	cgcgggcccg	gttggggggc	gcgagaactc	gaacgcgcgc	agatccggcg	5160
ccctcgtct	cgccgcgcga	tcgaaatcga	gcggcgccctg	cggaatctcg	atgccggcgt	5220
cgatcgccctc	cctggcggt	gccgacaggt	gcagatcgca	cgaggccgcg	tttgcgaaac	5280
actctgcggc	ggcggttcgtg	acgttggttcg	tcgcctccgc	ctcaccgcgc	ttacggcgca	5340
ggatctgcgc	gtctgctcagg	ttgttgcgga	tgagcccggt	cgtcacggga	aaccggacgc	5400
cgatcgcca	cggcgtcgcc	gtgccgagcg	tgaagacgt	gttggtctcg	atccggaagt	5460
cggtggacgc	gttggtcttcg	atcgcttcgt	cgccccattc	gttcagggtg	cagatcacgt	5520
tggtccgcac	gatgccgcgc	tggtgatcga	tctgcttctc	gccgtcgcgt	gcgtaggccg	5580
tgatccccgc	ggccaggccg	atggcgatgc	ccctgaacga	atcgacgatg	acgttccggt	5640
cgatgagcgt	gtcctgctgag	ttcgcccagg	cgaggatcgc	cgggccggat	cgccagccgc	5700
ccgattgcgg	gccgcgaatc	cgcaggaaga	cggtgtcgcg	gatgaccac	ccccgggccc	5760
ccaggatgtc	cacgccgtcg	gtgtaatcgg	acggcgcgct	cgctcgtgtac	tcgaagcgcg	5820
agcacgccac	gaggccatcg	tcgcggaagc	gtccgtcagt	gccgacgctg	cctttgagca	5880
attgctgtcc	cgcatcgatg	agctggacgt	tgtagaccgt	ggctcgggac	gcgccaagct	5940
cgcccgccac	ctgaatcgcg	tggtaccga	cgtagccgac	cgtagatcc	gcaatcgtca	6000
cgtagtggtc	gccaaccgag	agcgccacgc	cgatcgcgct	gccggtcatg	ccgcgcgcgc	6060
gaatgacgac	gcggtcccga	tcgcccggagc	ggctccggag	cacggtgccg	gggacgttga	6120
tgctggccat	ccggtcgagc	cggtactcgc	cggtctcgag	cagaatcgtg	gtgcgcggct	6180
tgacgcgttc	gagcgccctgc	agcagctcgg	acgtccgcgc	cacggtgacg	acctgtttcg	6240
tcgcccgcgc	gcgccatccg	cagaagtgcg	gcaccggcgc	ggcgccctgg	ccggacggca	6300
gaagcgccac	gagtgccgcg	gccacgatga	ggcgagtcac	gccgccaccg	tgtcgacgtc	6360
gtgttcgcgc	aaccagagct	cgaggatgag	caggatttga	aggagtagc	cgcgcgactg	6420
tctcccgtga	cgcggtgctg	cgatgatggc	gcgcaacgtc	tcggggcgca	ggaacatccg	6480
cgtgcgcgcc	tcggggggcga	acagctcgcg	ctcgacggct	tcgagcagtc	gcgtctgcag	6540
ccagatcgat	aagtcgtgat	agtgaacgta	cccaggcgcg	ttcaggcgct	tcagcagggt	6600
gttgaccttg	tcgaacacct	tctccgccag	cgaccggcg	tcgccgcgcg	cgccggtgtt	6660
cgagttccgc	accttgagca	gcgcccgggt	gccggcctgc	gtgatcgcg	gatgaagcgc	6720
cgtgtcgtcg	cgccatgccg	gatcgccgct	cagcaggacc	tcgaggaact	cgcatccat	6780
gaacggcagc	cgacagtcga	cgacgggtgcg	gaagagatcg	agcgaggga	tcgtgaagcg	6840
ccgatgatgc	tcacgaagat	agagatagct	gctgagctcc	gcgggtgaga	gctccacggt	6900
gtcgaggagc	gcgtgcagcg	agtcggccgc	gccccgtccg	gagaggcgcg	cagcgtctgc	6960
tgtgaagagc	gtgtccagcg	gcaggccgct	gctgatgtag	ttggcgcgcg	acgtcatgta	7020
gtcgatgaac	tcgcccgcgc	ccgtcatcgc	gaacacgtgg	ccgtcgggtg	ggagcggcca	7080
cgccaggctc	gtcttggcga	gctctccgc	gtggccgcgc	agcaggacct	cgataccgga	7140
gtcccgcagg	aaatcgatcg	ccagcatctc	ggtagctccg	tggtcagat	acatgccatc	7200
ggtgagcgag	accatgcgcg	cctgggttcg	cagaaagtcc	ttgaggtacc	ggtcgtcgag	7260
ctcgaagaag	cggtgggtcg	taccgggtgag	cttcgacagc	tggtgcgcga	tcacctgatc	7320
ggcgccagccg	gcgacgcccc	gcgtgtacgt	gatgagcccg	ccaacaccgt	tcaccgcact	7380
cagcacgggc	cggtgttcga	ggccgcgcga	gagcgacagc	ccgaagcgat	gctgcgccga	7440
gagcgacttc	tcgaccgcgc	gcctgaacgt	cgctgcacac	gcgtggacgt	acgcgtcctt	7500
cgctcgtagc	ggtccgcggg	agagatcggc	cggtctgaaa	taccgggaca	cgctcgcggc	7560
gcctgacggc	cagtcgtacg	cgagcacggc	gccgggggtcg	aggagcgtga	cgcccgccgc	7620
caggggtctt	tcgcccagga	cgaagcccag	cgtagacgtaa	tcggcgcccg	cccggagggt	7680
gagagccggt	ttgcgcgcgc	aggtcctgag	cacggaagcc	agatccgatc	cgaacgcag	7740
tcggcgccgc	tcgaccgcgc	agtagatcgg	atacgagccg	aacggatccg	ttgcaagcag	7800
caactcgtct	cgcgccggat	cgatcacggc	gagcgaaaac	gcgcccgtcg	gccggggcgc	7860
catggcaggc	cccagccggt	cgtacaccaa	agccaggaga	tcggcgcgcg	tcccgttcgc	7920
cgaaccaaac	tcgctccgga	gcgcccgttc	gttgtagaga	tcgccgcgaa	agatcaccag	7980
cggccggctg	acgtcgcgcc	cgcccgccac	ctcgtcgagg	aagccgagcg	tgccgagccc	8040
gaccgcccac	cggctccgacg	gatctcgata	gacgaccgtg	cggcgcggcc	tgcccgactt	8100
cgccgctgac	gcggcgctcg	cgccgctgaa	ggcggtagat	gccggctcat	cgcgcaaac	8160
gacgcccagg	aggccgttca	tcagacgacc	ccgcccggcag	acacgcgcga	cggaacaaga	8220
ggtgtcgtct	gagcgcccg	aaccggccgc	gcggcggtcg	ccgccaccat	ccggttgagc	8280
gcaatcgaca	agccgaggaa	gtgccagagg	atctcgggtg	actgaaagat	gacgaacggt	8340
ccgcccacgc	agaatgcaat	gagggcggtc	tgaggcgcca	gcgcgtacgg	cgccagctcg	8400
gccaggggccg	gatcccgtt	cgagaccggg	cggaacctgc	ggcaggcgat	gagcgccagg	8460
aggagagtga	gcaggaacgc	cgcgaggccc	acgaacccaa	gctcagccac	gagaccgaac	8520
cacgcgctgt	gaaccgagcg	cccttgccc	caggcgccgt	ccgacgtgtc	gtagtcgttg	8580
taggcggccg	tgaaggcatt	gtgaccgacg	ccgtaaatcg	gccggtcatc	ggccatgcgg	8640
agcgcgacc	tccagaagt	cagccggctc	tgggccgatc	gctgatcgcc	aaccgcccgt	8700
tcgctcgcgt	cgatgctcga	gtcggacgag	ctgatcgagc	tcagtcgatc	ccagaactct	8760
tccgtcatca	cgagcgacat	cgcgacgaag	accgcccgc	cgccacagag	gacgagcatc	8820
ttgcgcttcg	agtagaggaa	atagacgagg	ccaagcacgg	cgagcgtgag	gaagccgccc	8880
cgggagtacg	tgctgatgcc	gcgatagagc	acccgatga	gcagcacgcc	cagcgcccac	8940
ttggcccatt	tcctcgcttc	cgtctgcac	agcgtgacga	tgagcgccac	gagcatgaac	9000

atgccgacgg	cgacgccgtt	gttgtcaccg	agcgagtaga	cgtcgttgaa	gttcttgatg	9060
cggggaatga	acaggagctg	accccagccc	tgcttgccgg	cctcgaaccc	gagggagagc	9120
gagatcacga	ggaagacgag	ccgcaaccgc	ttgactcccg	tcgtgaggac	tgcgagcagg	9180
tacgtcatca	cgggtgactt	catgaagtcg	atcatgtagc	cccacgcgaa	gtcctggtag	9240
ggcgacgcca	gcgtcgacga	caggctgtgg	agcccgaaga	aggcgaggag	cgccagccgc	9300
gcgtcgatcc	gcagcgccctg	cccggacagg	aacgcgatcg	cgagcgtgta	catgcccgcc	9360
aggagcgaga	tggttatgtg	gatgagaaag	tcgctccaca	cccagagctc	cggccggaag	9420
tacgcgatga	acaggttagaa	cagaatcgcg	tagaacggcc	cgcgacgcgc	gtggaaggcg	9480
ccgaaggcga	ggaggggcag	gacgaaagcg	gctcgaagca	ttgccgtgtg	ttacctgtta	9540
tcgggtcccg	tggtccagag	cggctgcgtc	cgggtctcct	ctacattgaa	caccttgtag	9600
agggccggaa	tcgaggtgaa	catgagcacc	acgaaaagaa	tcgccgacac	caccatgtag	9660
gcgaagaacc	cccgtgctt	gtagagcttc	tccggttctt	gcacggggct	gttcggcagc	9720
atcccgatgt	gcaggtagta	ggcgaaacatg	cccgcgccca	cggcgcgcaa	caggatcagc	9780
tcgaggtgat	agcggacgat	gaagatgccg	ccgaagaggg	cgacgacggg	ggcgtagaac	9840
accatgctca	cgagcagccg	ctcctcgttg	tagaaggcaa	acgacttcgc	gtacgaggcc	9900
gcgaccgacg	cgttgccgat	gtgccggaac	tccgcgtacc	gcttcgtcgc	catgaagaac	9960
gcgcccacca	tccagtacga	gatcacgagc	gacatcgccg	gcacgcgatc	ctgaatcagc	10020
gggaaccagc	cgaggaggag	ccggacggcg	ttgttcgccg	actcgctcag	cacgtcgagg	10080
tacggccact	ccttcgtgcg	aattggcgccg	acgttgtagc	tgacgccgag	caccagagc	10140
gccagggccg	acagcgcgaa	gtatctgttc	acgagcagcg	cgagcacgaa	tccggccacg	10200
cccacgagga	tccactcggg	gtagcccgcg	gccggcttga	tcttcccggg	gggcaccggc	10260
cgggtgccgtt	tctccggatg	cagcagatct	ctcgggccgt	ccaggagctc	gttcagcacg	10320
tagttgctcg	aggcgatgag	gcaggtcgcg	cgagcgcgga	gcgcgagcgg	cgggacggcg	10380
gtccagccga	agaagctgcg	ttcgtagaag	aacgccagca	gcacgccgag	cagcatgaac	10440
gcgttcttga	accagtgatc	gatccgcgcg	atctggacgt	acggccaaat	ccgcgaaccg	10500
gagctagcgg	ctgacgacat	cgttgagctc	ctcgagccgc	cggatcgcg	accatccgcg	10560
gccggctgtg	acgccgtggc	ggccggcgat	gagggcgag	gagagtcctg	cggccatcgc	10620
tccttctctcg	tccacgtccg	gtcgatcccc	gacatcacag	acctcccag	gtccgagtga	10680
ccacatctca	cacgcccacat	ggaagccgcg	cggttgccgg	ttcaggggcg	tcacgttgct	10740
ggcggtcgtg	cagagcgcca	gcgaaaaatg	ctgcgacgtg	ccgagcgccg	cgagcttgct	10800
ctcgggcgcg	tagtccgaca	gcacgccgag	cctgagcccc	gccgcgcgaa	gcgagtcgag	10860
cgcgccgagc	aaccccgggc	gccggcacca	ccgcagatac	ttgagcgggc	ggcgaccatc	10920
ccattcgttc	accgtcgccg	ccacgacctc	gttgctcagc	ccgagcgccg	gcgcggtgcg	10980
cgcgatctgg	cgcgcgcgga	cggtctcgtc	ggccgcgccc	aggcgccgga	gatcctcatg	11040
cgcccgccgg	tactcccga	cgatgcgcgc	cgtgtccact	ccgcgcgccc	accggaacgt	11100
cacgagcggc	gtggcgcgga	gctccgctgc	catggcgccg	cgcaacggac	cctgccggta	11160
cagcgtgccg	tccacgtcga	acagcacggc	cttgaccccg	cgcaacggct	tactctgcg	11220
aggccacga	tccgcagaat	cgccgacacc	tttgacgcgt	cgcgaaaggaa	cacgtcacga	11280
accacacagg	cagccagccc	accgcgcgtg	ccgcgcgccc	ccgcgcgccc	tgccgtccgc	11340
cagctcccgc	gccaggcgcg	cgtgtcgatc	gcggcgccgc	cgtagcagag	cagaatcggg	11400
acgagcgcca	ggtgataccg	gggatgtccg	aagacgatcg	cgtgaatgcc	gctcatgaac	11460
acgatcaggc	tcaggagcag	caggtgcgcg	cgccagttga	ccggccgcgc	gaacgagatg	11520
ccgatgacgc	cgagcaccat	gaccgcgcgc	tacgatccca	tgatggacac	cgccccggcg	11580
aagaacagcc	accgcggcgg	acggtagaac	ccatttcgga	cgccggcgat	gaagtgcgcg	11640
tccagccccc	agaagtctgc	gaacttcagc	accgaacggc	gaagcgtcgt	tcccggatgg	11700
gccttcatgt	agtcgagcgc	ctggcgctgc	gcccacttct	cttctcgtgc	ctccgtccag	11760
cgggtgccgt	ccggcgccgc	aggcgccatc	gtgcgcgacc	attccttgct	gcccgtgagc	11820
gagatcgcg	cccacatgcg	atcctcgggc	gtgtacgcgt	agttgcccat	catcaggttc	11880
agcccgccga	gcgtgtccac	gagcgttga	gagcgttga	ggagcgtgtt	gcgcacgctc	11940
cacgggccga	cgatcgccgc	gtatccggcg	aagagaagca	cggccacgcg	cagacgcgcg	12000
cgaagcgcca	tgcccagacc	gacgagcgcc	gcggccatga	ggatcgcgac	gtacggccac	12060
atgatcgagc	gaacgagcgc	ggacgcgcgc	agcgcgacgc	ccgtcgcgag	cgcgatccac	12120
gccgagtggc	gcccgtatgg	gcgatcgatg	agcgtcagac	atcccaagag	caacaaaaga	12180
aggagtgtga	tgaagagcgt	ctcggaacgc	accagcacgc	cggagaacag	cagcgaggga	12240
tagaacgcga	acgccgccgc	ggcaatgagg	ccggcgccgtg	cgctgaacaa	cctgcgcccg	12300
atgagataca	cgagccagac	accgagcagg	ctcaacggaa	cctgcgccag	ccggaccgcc	12360
gtcagcgatt	ccgagcccgt	gaccgccccg	acgcccgcga	cgaacgcggg	aaagagcggc	12420
gcccggatgg	acgtcgggtc	ccccggcccc	cacgcgaacc	cgttcccggc	gacgacgttt	12480
ctggcaagcg	tcgcgtagt	ctgctcgtcg	cgaatctcga	gcgcgacgtc	cttgaacccc	12540
caaatacagg	cgagccggag	gaccagggcc	agcgtgagga	tgacgagaat	tgcccggcgc	12600
tcccaatggg	tgcgctcggc	ctcgctcggc	ccggggcgga	cgattggcca	ggggggcgac	12660
gatgggagt	ggggagccctg	catcattcgc	taaaacaagc	agttacggtc	aaacctgtgc	12720
caccggggaa	tcagcaaaga	ctctgccaac	gggtcccgccg	gcggcgacgc	ccggcgccgcg	12780
ttttccggcc	ggaagtcggc	ccccaaaggc	acccggtcgg	gcacagccca	gtagaaatct	12840
cattctggga	tgacttttgt	ccatgatcca	tgaccatttc	gaggggcctt	cgtgacgccc	12900
tggggatggc	ataagcttgg	cacctgcgat	ggcgacgtgt	cgatgatcag	cgcccggtgag	12960
gagctccaga	aggcctatcg	ggacgacccg	gtcgcccggtg	aatacgtggc	gcggcggttc	13020
cagtcaccgc	tgggggcgct	cctccattcg	cgtcagatcg	gggtcggtcg	agagctcgtc	13080
cgcgcgcagg	gcacccggcg	cgcggcgagg	atcgcacccg	gtccagcgcg	gctgaccgtg	13140

gacatcgcg cgtctctcga ccgctgacg ctcgtcgacg cgagcgccca gatgcttcac 13200  
gaagcgcgcg agcggtcgcg cgaccggacg ctcgcccgcg gggccaatct cgtgcaggcc 13260  
gacgctgttc gtctgcccgt cagcgggccg ttcgatctcg tctacacctt caggctcgtg 13320  
cgccatttcg agcgcgccga tcggctgcga ctctaccgcc agattgacgac gattctcggg 13380  
cccggcggtt ggctcgtctt cgatgcggtg aatgagattg tgtctgcgcc gcttcgcgcc 13440  
aacgcgaagc ccggcgagta cgagcactac gatgcgctgc tcacaccggc cgggctacgc 13500  
gacgagctgg ccgagggcgg gtccacgatg gtgtcgctca ccggtgcgca tcgcccgttc 13560  
cccacgctca tgaagtgccg gatgtatctc gcgcgaggt caacgggtgt cgcgcgcgcg 13620  
gcgatggagg tgatcgatcg gctcggcgcc gaaccgctcg agtggatcgt cgtatgcgcg 13680  
cgcggtgtac ctactggacg ggtacctggg atccgatcaa agaagcgatt tcgaaggaa 13740  
tgaacgccct ccgacccggg tcgcccggcg acgcgacggt cgtgtcgttc tcgcccggcc 13800  
agcgagccg tctcgtgcg aaagatcgcg tgctcgtgtt gtccggccgc gcctggatcg 13860  
cccttcgggg cgcgccgccc gtgtcgcgac cgcgcgcgca cgtgacgcac gtcttcggcg 13920  
gcggtgtctc ctggcatctc ctgagggcg tgggcgcgcg gccggttctg ctcaccgccc 13980  
tcaaggaa tggcgctgcc gatctgccc gggcgatcgc gcacgtggcc gtcgaagtgc 14040  
gccgcgctcg cgacacgtgg atcgggccg gttatctcc cggccgctg cacctcattc 14100  
atcccgcggt ggacctcgac ttgtatcgcc cgtgtccggt ctcgtcggag agattcacgc 14160  
tgctcttcgc cagcacgccg tccgatccgg ctgacgtgct ggtgccgtgg attccgctcg 14220  
tcgtcgagct ggcgcggctc cgtccggaga ctcgacgtgt tccgccaccg aacttcacgc 14280  
gcgacgtcga cgcctccagg cgcgcgctcg acgcccctcg gccgcgcgcc attccgctcg 14340  
tctcgtacga ggacgcacc gacatgcgcg cctgttacgc gcggcgccac gccaccgtcg 14400  
tgtgtctcgc gggcgccacc ggcaaggcct gtcgaattt cgtgtcgcag ggattggcca 14460  
tgggacggcc gtgtctctcg acgcgcgaga acggcatcgc ggagatctc gatcgatcgg 14520  
gcgcccggcat catcgccgcg cgcgatgccg cgtcgtcgc cgcgagcctg gacacgctcc 14580  
gcgcccgtg ggcgggctgc tacgacgcgc tgtacgtgga ggtggcatcc gccccgctg 14640  
tcggcgctt ccgcccgcg tcttgacgaa gggactcgcc ggacgctgc tcgggtgggc 14700  
gcgagcacgc gaggcgccg tcttgacgaa gcccatcga tcgcagatga agcgcgacgc 14760  
tcagaaggcg aacgatttcg ccgactacgt gctcggcgcc cgcgtaacc ggtatccat 14820  
gcggcgcaat cgccgcagg tccatcgggc gtcgacgcag ggcgtcctcg gggccctgcg 14880  
cgtgcccgtg gccagcatcg gcgacgtcgc gtcgacgcag ggcgtcctcg gggccctgcg 14940  
cgcgagctac ccgagggcga cgatcgactt cctgacgtcg gacgcgaccg tggccgtcat 15000  
ccggcaccac ccgcgcatca acaccgtcta cagcgtccag gatctgtcgc gcctgtccga 15060  
gtacgactgg atcatcaacc tccagaacct cgtgcccgcc gcgagctttc tcagaacgtc 15120  
cggtctctcg taccgcgaga gtcacgcggc gtcctcgagc acgaacatcc tctactgcat 15180  
caccggccgt caccctcgaga acggctcgga gtccacgcgg cgcagacag cccgctaccc 15240  
ctgtgagatc gaggagcatt tccctacggc gtcctcctcg tacgacagc cccgctaccc 15300  
ggtgaccgag attcacgtgg acgcacaggc acgagcgcg gcacgatcga agttcgctt 15360  
cccggagaa cgccaggtgc tggcgatctt cctcggttc aactcggtcg ggtgcggcg 15420  
cgatgagggc tttcgacgt actccatcga ctacgtgaag cgcctcgtgc gtcactcga 15480  
agatcggtt ggatcgcca tcatcgggca gtcccagggt gatctcgtg acaagacgtc 15540  
acagtaccgc gagatgctcg gccgccatcc ggacgtgatc ctcatcagtt gcgactcgag 15600  
gctcgacgag ctggtggccc tcatggactc gccggtcatc gggctgtacg tcaacgacgg 15660  
ccccgtccat ctggcgatgg ccgcgaggt gtcgacgttc gtggccatca acagcacgc 15720  
gacgttcagg atgcacccga cgtcggcgga cgagcgcttc tgcagctgac gcgacccggc 15780  
gccgtgcttc acctactcgt ggcgatggaa gttctctgc tcgacgtgac gattccgct 15840  
gaccagatcg cattactgtc acaacgaggc gttcgtgtt ggcgtggacc gattccgct 15900  
ggccgcaatc gaccgcggg cggcgcggt gctcaagggt agtcccgat caggacctca 15960  
ttcgcgcgca gtttgaagag tcgatcgaga cgaagcgcaa agcgtctcg ctcgtcgatg 16020  
cgggtggcga gtcggcgcg ctcctcatcg agcgctcga ggcgggacac aagatcctcg 16080  
tgtgcccga cggcggtcg gccgccgact ccgtggcgct cagaccgac acgtcgatcc 16140  
gcttcgagaa ggagcgccg gcgttgccg agaccgtgtc agccggcagg tgagtgcctc 16200  
tcaccgctg gccgaacgac tacgctacg acgacgtccg gcaactcgcc caacatctc 16260  
ggagcgccg gcgacgtgt catcggtac atgaaacgaa tcgagctcac cggacgcggc 16320  
gccgccatg acgagggcga aaagaaacgc tgccaggtca tgcgtccgtc agatcagacg 16380  
ggcggaaga tgaagagtct cgccggcgct atccacatct gggcgaggt gatcgagaa 16440  
tcgaggatcc aggaagtaca cattaccgtc acgtgcgac ttcgaaagtc gaccagagt 16500  
gcgtcagct gagctggcg tagacgtcga cgtgagggc gcggcgagg tgctatcct 16560  
gctcggggc gatgaggtcg gcggcgcggt cctcggcgct gtcgcccgtg acgagaagc 16620  
cgaacaacag ttgcagtcga ttcgctgac cgcccggat ggtggagccg atcacgggca 16680  
cgggtctccc gtcttcacg acttcacgca gcggcagccc ctcgtgctt gacggcagca 16740  
cgcgggccga ctgggctcg agcaccgaga gcgctcctg gcggaagccg agaaagcgca 16800  
ggaaggcgtc ggcggcctg agcagcacgc ggcgctcctg cgcctgctcg tgctccttgc 16860  
cgcatcctc gagcgcgagc tcggcagaga ggcgctcag caggccgccc gacatcagca 16920  
cctcggtcgt gtcgctccc accagccagc ctttgtgtgg cgcaagattc gcgagcatga 17040  
gcctggtcgc ccgatttgc gtacgtcgc cggccgtac cgcgcttgc cgttcgccc 17100  
ggacgacgaa gcggtcggtc tcgagggcgg gaatcgcggt gaccggcacc ttgtcccggc 17160  
ccacgaact ctcccgatcg acccggttcg cgtgtagcg cgcacacgtg acgatgtgt 17220  
gtccctgctg cgccgctcg atcatcacg cgttcggtat cggttcgatg tgaagtgca 17280  
gcggcggtt cttcaggacc caggcgatct cctcgggatc

cgggccacgcg	cgcgcgcgcg	gcctgcagcg	cggtgtctcat	gaatccgtac	acgacaggat	17340
tggtgcacgtg	gaccaggggc	cgcccgcagc	gacccaatcc	cggaagcatg	cgcgcacacg	17400
cgggcgagctg	ttgcaggccg	cgggcccgcg	cgcgcttcag	gtcatacatc	cgccgcgcga	17460
cgccctccgc	gtcgagcgcg	gcgcccgcag	gaccttcgcc	tgtaatccat	gccgtgcacg	17520
ggacgcccgcg	gcgtttggcg	gcggccgcga	accggatcgc	cagcaccgcc	gcgcccgcga	17580
ccagatgtga	gacgaggatc	tcgtgcaggc	ggggggactg	attcatgggtg	gcgcgcgatg	17640
cgggccgggccc	gccagcctac	gttactcgcg	tcgcagggtcc	gcccgcgcgg	tagagcgagc	17700
gcagcaggct	gcgaagtccc	gacgggagga	agcacacggc	gtaccgcgtg	agatcggtgc	17760
gccgcacgcg	caggcccgcg	tgaccgaacc	gtttccgcgc	cgctgaaac	tcgccgcggt	17820
cgtactcgag	gtggccgagc	gcgagcagga	gacgatggcg	atgatcgctg	gcggcgcgat	17880
gcgcccactg	atcgctcccc	cgcgccagga	accgctccag	cgctctgacc	atgccttcga	17940
gcatcacgat	ggacgacttc	gacatgtttg	tgctgtgcag	cctgtaatgc	gtcaggggtt	18000
ccgggatgaa	accgaccggg	cccttcgcgc	agaggcgag	ccagaggtcc	cagtcctcga	18060
cgccgtggat	gccggtcgcg	aacatctcct	gccggagcac	ggcggtttg	accatgaccg	18120
acgaagcgggt	cacggcggtt	cgagggacga	ggtcggggag	acagcgcgcc	tcgagcggcg	18180
gaagatcgag	cgtgcgcagg	acctcgccgc	cctcgctgat	gacggcgcg	cccgtatgaa	18240
tgaagctcca	gtcggcggtc	tgctcggaaga	gatcgacctg	ccggcgaaac	ttggccggat	18300
cccacgagtc	gtcggcgctc	agcagcgcca	cgaattcccc	tcgcgcgcgc	gcgatggcgc	18360
gattgcgggc	gccggcaagg	ccccggttgg	cctgagagat	caccgtgatg	gcgcccgcga	18420
atctcgccag	gatgtctgcc	gtgcgctcgc	tcgatccgctc	gtcgacgacg	atcacttcga	18480
cgggctgata	ggtctgggag	agcgcgagct	gcacggccct	ggcgatgtag	cgctcgccat	18540
tgaacaccgcg	gacgatcacc	gacacgaggg	cggggatata	ggtctgggag	gccggcgag	18600
gctctcgagc	ggcttcagcg	gccatggccc	tcgtcacttg	cgctcgagag	gaacagcatg	18660
acgtttctgc	aaagccacgc	agacaacggc	cgctcttcca	tcgcctcccg	cagcgtctgc	18720
gatgccgggc	cgaggtagct	gaaccgcgct	gccgcgaaga	tcttctgcca	gtactcgagc	18780
ggctgttctg	tgatgtggcc	cattccgcgc	tggcccggct	gcgcgcgct	gaagagcacc	18840
agcgcgccgc	tggtcgtgag	cacggcgacg	agcttctcgc	caagcgcgcg	cggaagggtg	18900
tcggcgacct	cgaggtctcat	ggagaggtcc	gttgctgcgc	cgcccggtga	caccgcgctg	18960
agatcggaat	agttcagcgc	gacgcgcgag	ttttccgcgc	cgtaccgtcg	cgctactcgc	19020
gatcgctcgt	agccggccgc	cttcacgcct	cgtgcgagga	acgcgtgcac	gaagtgaccg	19080
gtgccgcagc	cgaagtctgt	catcgatcgc	atgccgggtc	gcaacgccag	gaccccatca	19140
acgagcacgcg	gcgcccgcctc	ggcgcgacgc	tgcccgaaca	gatccatgac	ggccgcggtg	19200
tacgtccatc	gctcgaagcc	cagccgttcg	cgatccggg	acatcttcag	tcggccctgg	19260
caacgccccg	accagtagag	ccgcttgagc	tgggtcacga	tcattgggttc	gtcccgggag	19320
cgacgcgaac	gcgcccgggc	gcatgcagcg	acgcccagaa	ctccacatcg	gccacgcgcg	19380
cccacgatcg	atgttccttg	accgccaagc	tcgcaagctg	ccgcgacggc	acgaggaacg	19440
tctgggtcgc	gttgtgaaac	acgtgcaccg	acaggtagta	gtggccgcga	acgagatgcg	19500
cgcggaatg	gaggtcgagc	gagaactcgc	ccatgaacgc	ggacgcgtcc	ccgagatccc	19560
gccggtggat	gttgccgtcg	tagacgacga	cgtgatcggt	ggagcgttgc	accatgaagc	19620
cgaacgtgac	gtcgtccagc	ggctcgggtac	ttctcgcggt	gacctggagg	ttcagctgct	19680
tcggcgcggt	cacttcgctg	tgctgcgcgc	acgactcgtc	gccgacgaag	agcgagccgc	19740
acacgatctc	gacggggctc	ccgacactcg	gctgttccga	tttcttgga	gcgcgagcgt	19800
acgcgtcgag	gacttcgatg	gccgtgcgcg	cgcccttcac	cgagccggcc	agatacaccc	19860
catgctcgca	gaggttcgcg	accgcctgca	gggtgtgcga	gacgaggacg	atcgcgatgc	19920
cctggcgctt	gaactcctgc	atgcgatcga	agcacttctg	ctggaacgct	atgtcgccga	19980
cgctcaacac	ctcgtcgatg	atcagcacgt	cgggccgcag	gtgcgcggcg	atcgagaagc	20040
cgagccgcgc	gttcatgccc	gacgagtagc	gcttgacggg	cgtgtccacg	aagtccggtg	20100
gcggcgcaaa	ctcgacgatc	tcgtcgaacc	gctcggcgat	ctcgaccgtg	ttcatcccca	20160
tgatcgcaac	ctggaggaac	agcttctcgc	gcccggtgag	gtcgggatgg	aaccccgcg	20220
cgacttcgat	gagcgcgccg	gcgcccgcgc	ggacgatcga	gcggccgggtc	gtcggcttca	20280
ggatcttctg	cagcaccttg	aggggtggcg	acttgcccgc	cccgtttgcg	ccgatgatgc	20340
cgagcgccctc	gccgcggtgg	acttcgaacg	agacgtttct	gaccgcccag	aactcctggt	20400
cgtccagcgc	gtcggcgggg	cgcggttgaa	acagcccgcg	cacgagcgcc	ggcacgaggt	20460
cccgcaggct	gtcatgacgc	tcgcccgcgc	ggaacttctt	ggagacgttg	tcgaagacga	20520
cgggggccgt	cacgtcagac	gttctcgcgc	aattcgaact	cggagcgatg	gaagaacagc	20580
cagccgaaga	cgagcgtcac	gaccgagacg	atcgccgtga	cgcccgatg	cacggacggc	20640
ggctggttca	ggagcagcac	ggcgcggaag	ccgtcgatga	tcggcgctcat	cggttgagc	20700
tgagggggcc	actggagctt	gcccggccac	atttcgaccg	ggtagaccgt	cgacgacgcg	20760
aacatccaga	cgtcgatgac	gacctgaag	aggtatttca	cgtagcggtg	gaagaggttc	20820
gccatcgaca	ggatcagggc	gatggccgtg	gtgaatatca	cgagcacggc	gagcacggcc	20880
ggcagccaga	ggatgttcgc	tcgacccggc	acgtggtagt	agatcatgag	cgcgacgagc	20940
accacggagc	cgatcgcgaa	gtccacggcg	ctcacgatca	cgccggagaa	cggaagatc	21000
tccttgggaa	agtaacacct	cgagacgaga	ttcgtgttcc	cgtcagggga	ggtcacggag	21060
aagcggagcg	ctagagagaa	gaagttccag	accacagggc	cgcagaacgc	gaacaccgga	21120
tacgcgaccg	gcgtgtcgat	ggccgcgacg	cgcatgaaga	tcacggagaa	gacagccgtg	21180
ttgatgagcg	gcatgaacac	ggcccacccg	aagcccatca	ccgactgctt	gtagcgacgc	21240
agcaggtcgc	gccgcgtcat	ctgcatcagc	agctcgcgat	actcgaactg	ctcggcgagc	21300
atctcacggc	aatcggccag	gaaggcgcg	atcgtcagcg	gtgtcagtcg	cgccgtcgtg	21360
ccggcgggat	catgagggca	ggtagcgcgc	gagctgcccg	aagagcaact	gcacgaactc	21420

ggctcgcgcgc	cgctcggcgg	cgacgtcgcc	gttttcgctc	gacggcacgg	ggctcgtcac	21480
gcggcagagc	gcggtcgcg	atcggttcag	ccgcatcgcg	tcgtacacca	tgtagatcct	21540
gctccagtag	tcgctggcga	tcacgcggcc	gtggctctgg	taccagtaca	gcacgatttg	21600
ccgctcgatc	cccttctgga	tgacgtatcg	gttgatctcg	atgggctcgg	tgccggcccg	21660
caccttgacc	agcgcgcgcc	cggtggatac	gggctcccag	ccggcgcccc	gcaggcagtt	21720
catcgcgag	tggtatcggt	cgcttccgcg	ctggctctgg	tagtagccga	tgtagagcga	21780
cacgtagggc	tcgcgcggcg	ccaggtagat	ccggttgatg	tactcgtcca	cgccgagcac	21840
gcctcatgac	tccttggtga	agggcgcgct	gtcgcgtccg	gtccaccgct	ccatctggaa	21900
cgggacgttc	cgagcgggt	ggcggaagcg	gacctgctcg	gcgcgcacat	agcgcgcggc	21960
aaacgccgcc	gtgccgagga	aacacacgct	gagaatgatc	agtcgtctgg	tcataaatct	22020
ggactcacac	cgtggccggc	gcggtcaccc	ctcgggcgcg	aggctggaag	cgccgcagga	22080
cccagaccac	gacgaacaac	atgacgaacg	cggaagaaa	gacgagccag	cccagagaacg	22140
tgtggaagaa	ccccctgcgc	gcggcctcgc	cgtagtagtg	cgccgccacg	cccgtgcccg	22200
ccacacgcgc	gcccgttcg	ataatcgcg	tcgggatcgt	cgagagcgca	atcgccaggc	22260
gcatgccgac	gcgcggatcc	gtgaagtagc	cgtagacgat	gcccagcgtg	agcagcgaga	22320
tgagcgaccg	gatcccgtg	cacgcctcgg	ccacttcgag	cgctcgtgtg	gcgagcacga	22380
tcacgttgcc	ttcgcgcagc	accgggatgc	tgagcgcgga	cagcacgccc	tcgcccacgc	22440
gagacgccag	gagctgcagc	ggaaatgcga	tctgattgaa	gatgatcgac	gggatcggga	22500
tcatacgcac	caggaacgcg	agcggaacacg	cgagcaccca	gagcttgcgc	cagccgagca	22560
ggaacacgat	cgcgccggta	agcgaggcta	gtatcgacac	gcgcgtcaga	aagagctcgg	22620
cgccgagcaa	ccccgcgacg	agcagcgcg	gcccggcgag	cacgatgacg	agaccgagcg	22680
cgctcggcgt	gtccggcagc	gccgagagac	gcgcccggcg	ctcccaggca	aagtaggccc	22740
cgagcggtag	gatcagaatc	ccgtgggagt	agttgtcgtc	gttgatccag	tcctgcgcca	22800
accgacgat	cacctgcccc	ttaaaggagca	ggaaactcac	ggcggcgata	ccgagcgga	22860
ccagtctggc	ctgaagcaca	ctcatagcga	tcgggacgaa	tatagcaaag	ggctctccc	22920
ggcagaactt	ggcgccgcaa	acacgtaacc	atcacgctca	caggaagtta	gcggctaact	22980
gacgggattg	gccggctgga	ggggcctcgt	gaatctaccc	aatggatgac	aaatatccct	23040
atcaactcct	ccttttgact	gcgtagcacg	gaagatcac	ggcctcttca	gatgcaggac	23100
aaatgtcatc	tgtaggaag	agtaatgtct	ctggccacag	cgccggtcaa	gtgaaatccc	23160
ccgacagaat	cgcgtagcac	cgcgactggc	atccggcgag	aaaagtcgcc	tcgtcggccc	23220
gaaaactgca	ttacaggaac	gccaggctcg	cagccggttc	gttttcgcgt	agtatttttg	23280
ggcattgcag	gcgctttcgt	ctgccggggg	gcggaaccgt	gcgaccgggc	cgatcacgcg	23340
tggagcagg	cagggacagg	catggatggt	tcctgtagag	agccgcgttc	gggagacggc	23400
cgggctcgcc	gggagctcgt	ccacctggcc	gtggaccgcg	cgggcgccca	atttctctgac	23460
gcgcggcgcc	taagttttga	cgcgctctcg	ctctcgtatc	gcggcctgac	tcgcccgcgc	23520
aacgcgctcg	cgaccgcct	gcagcgtcat	ggcgtcgatc	ctgacgtggt	ggtcggcatc	23580
tcctctgacc	ggtccatcga	gctcatcgtc	gggatcctcg	ccgtgctcaa	ggccggcggc	23640
gcctacgcgc	cgctcgatcc	tgcgtatccg	aaggaccgcc	tcgagtacat	ggtcagggac	23700
gcgcggcgcc	ggatcggtct	ccactcgca	cgctgagcgc	gtctgatccg	cgctcgatgg	23760
gtcgagctca	tctgcctcga	tgccgacgag	cccggggcgt	tgccggaaga	gggccatgcg	23820
ccggacgcgc	gcgtcgatct	cgaccacctc	gcgtacgtca	tctacacctc	gggatcgacc	23880
ggcgccccga	agggcgctcg	gatgccgcac	cgctcgttgg	cgaatctcat	cggtggcag	23940
gtcgagcgc	cgccgctgg	cgccggcgcg	cgacacgtcc	agttcacttc	gcccagcttc	24000
gacgtctcgt	tccaggaaat	ctttgcgacg	ctcgcgaccg	gcggcgagct	cggtcgtctg	24060
tccgaggaca	cgcgccggga	tccgcgcgcg	ctgctgcggg	tgctgcgcga	tcgatccgtc	24120
gcccggctgt	atctcccgtt	cgctcgcttg	cagcagcttg	ccgtcaccgg	caccgacgat	24180
ccggatctgc	cggcgttgcg	tgaggtcatc	acggccggcg	agcagctccg	ctcgacgccc	24240
gcgctcgtcc	agtggttcgc	gcggcatccg	cagtgcacgc	tccacaacca	ctacggcccc	24300
gccgaagcgc	atgtcgtcac	tgcgacacgc	ctgactagcc	cccccgcgga	atggccggcc	24360
ctgccgcgca	ttggcggtgc	gctgcccggc	gtcgagattc	tgatcgtcga	cgagcatcgc	24420
cagcccgctg	cgacagcga	cgtgggcgag	atcctcattg	gcggcggtg	cctcgtctgc	24480
ggctacttga	atcagccgca	gcgcacggcc	gaacggtttc	tctcgcaccc	gcttcgcccc	24540
gatgctcgcg	tgtatcgcac	ggcgatctc	ggcgtgtg	gcgacgatgg	cgcatcgag	24600
ttcgtcgggc	ggacggacga	ccaggtgaag	attcgaggct	tccgtgtgga	gcccggcgaa	24660
atcgagaccg	tcctccttca	gcatcccgg	gttcgggaga	ccgcccgtct	cgcgacgac	24720
gacgcgtcgg	gcagcaagag	gctggctcgc	tatgtgatcc	cgtagccga	tcaggcgcc	24780
accggagccg	aggcttcggc	acaggtctcc	ggatggcaga	ccgtgtggga	ggacacctac	24840
cgcgcgcgcg	cggaactcgac	ggatccggcg	cttgccgcga	ccggctggcg	gagcagctac	24900
acggggatgc	cctacaccgc	ggcgagatg	cgggagtg	ccgatgagac	cgctcgcgcg	24960
atccggtcgc	atgcgccacg	cgaggtgctc	gagatcggt	gcggcacggg	catggtgctc	25020
ttcgtctcgc	cgccccactg	cgctgtctat	agcgcgacgg	atgtctctcg	cgccgcgtc	25080
gaccacgtcc	gggcccacgc	gacggcctgc	ggcgccagcc	acgtcgagct	cctctcgcgc	25140
gaggcgcgcg	actttaccgg	cgctccggcg	gcccgtgtcg	atgccgttgt	gatgaactcg	25200
gtcgcgcagc	acttgccctg	cgtggactat	ctgacgcgcg	tcctcgccgg	cgccgcccgc	25260
accgtgcgtc	cgggcggtat	catcttcgtc	ggcgacgtgc	cgaacctgcg	cctgctcgag	25320
atgtttcacg	cgtcggtcga	gctcacgaaa	gcgcggcgcg	atcttccgat	tgccctcgtg	25380
cgcgagcgcg	tgacagctca	gatggcgctc	gagcgcgagc	tcgtgatcga	tccgatgttc	25440
ttcgacgccg	tcattggaccg	catccccggc	atctcgggag	cccaggtcca	gatccggcag	25500
ggtgccttcg	acacggagat	gaaccgggtc	aggtacgacg	tcgtgctcac	ggtggcgga	25560



cgcccgcccg	ggcgccgctc	gtccgtcgcg	ctcgactggc	gtacggagcc	ggtcgcccag	25620
cgtaagctcg	ggcaactgct	cagcgccgat	gcgcccggag	cgatcgagat	caggaacgtg	25680
accaacgcgc	ggctgctgca	gccggccgcg	tgcttcgatc	tcgtcaagcg	tgccgactgg	25740
accggcacgg	ccggcgagag	gcttgagagg	gccgcccggg	acacgcgcga	cgcgatcgac	25800
ccggaatcgt	ggcgccgctc	cgcgaacgcc	gccggctacg	acgccgagtt	cacctggctg	25860
gccgagatcg	gcgccatcgg	cttcgatgcg	cggttcgagc	gccggggcac	cgcgtcgctc	25920
gccgctcccg	tgactcgtca	cgcggttccc	gcccgggagc	cgcggggccat	tcagcagtac	25980
gcgacgaatc	cgttcgcgca	tgccgtggcg	cggcgcccca	cgcccgagct	tcggcgattc	26040
ctgcaggagc	gcgtgcccga	tcacctgggt	ccgtcgccct	tcgtcgtaact	cgatgcgatg	26100
ccgctcacgc	caagcgggaa	gatcgatcgg	cgcgcccctc	cggcgcccga	gagccgcccgt	26160
cccgatctgg	acgtggcggt	tgccaagccc	gccacggagc	tcgaacgcaa	gatcgcgag	26220
gtctggcaga	cgacgctcca	gatctcgctg	gtgggattgc	acgacaactt	cttcgatctg	26280
ggcgggcact	cgtcctgct	cgcgcaggcg	ttcgagcgta	ttcgcccgct	cgtgcccggc	26340
aagcagtggt	cgatgatcga	gatgttccag	taccgcagcg	tctattcgct	cgcggggttt	26400
ctcagtgagc	ggcgggacgc	cggcgccggg	cagcttgccg	gcgcgcagga	tcggcgacgc	26460
cgccagcgcg	agatgctgac	gcgtcagggg	cctgccctga	gcagagtcga	aggggtccgg	26520
taatgcagaa	cggcatcgcc	ctcgtcgggg	tgcccgcccg	ctttccgggc	gcgcggctga	26580
tcgaggagta	ctggcgcaac	ctcgtcgccg	gcacggagtc	gatcacggtc	ttcaccacgc	26640
aggagctggc	gacggccggg	atcagccaga	gcgatctgca	gaatcccagc	tacgtccgcg	26700
cgcgcgggct	cgtcccggat	ccggacctct	tcgacgccgc	gttcttcggc	atcagccagc	26760
gcgaagccga	gctgatggat	ccgcagcacc	ggctcttcc	cgaggcgctg	tggtccgccc	26820
tcgaacacgc	ggcgatcgcg	ccgcggtcgt	tccccggct	catcgccgtg	tggggcccga	26880
tgagcacggg	gatgacgaac	agcacgtacc	tgctgtcgaa	tctccacagc	cacgcgggat	26940
tgctccagcc	ggaggacgtg	ctgcccggcg	tgctgcccga	cgagaacgac	tacctcacga	27000
gcccgtcttc	gtacaagctg	cacctgcgcg	gaccgagcat	gaacgtgcag	accgcgtgct	27060
cgacgtcgct	cgtcgccggt	gtgcaggcgt	gtcagtcgct	gctcacgtgg	cagtgtgatg	27120
cggcgctcgc	cggcgccggt	tccgtgtcgt	tcccgcaca	ggaaggctac	gtctacgtcg	27180
agggcgccat	cgggtcgcca	gacgggcact	gccgcgcgtt	cgatgcgaac	gcccgaaggc	27240
cggctcttcag	caacggcgct	ggcatcgctg	tccgtcgaa	gctcgaggag	gcgatccgcc	27300
acgggcagac	gatctacgcg	gtgatccgcg	gatcgccgct	gaacaacgac	ggctcggcga	27360
aggtcagttt	cgcagccccg	agcgtggacg	gccaggcgga	ggctgctggc	atggcgagca	27420
gcgcccgcga	cgtggcgccg	cagacgatcg	actacgtcga	ggcgacccgc	accggcacgc	27480
cgtccggaga	tccgatcgag	gtggctgcgc	tcacgaaggc	ctttcgccgc	ggcgccgcca	27540
cggagaaagg	gttctgcgga	ctcgatcgg	tcaagagcaa	tttcgggcac	ctcgactcgg	27600
cggccggcgt	cgcggggcgt	atcaaaacgg	cgtcgcgcgt	ccatcaccag	cagattccgg	27660
cgacgtcgca	ctacgtcgag	cccaaccgca	agtgtgattt	cgccaacagt	ccgttcttcg	27720
tgaacgcgcg	gctgcgacca	tgcccgccgc	gtggccatcc	gcagcgccgc	gccgtgagct	27780
cgttcgggat	cggcggggacc	aacgcgcacg	tcgtgtcga	ggaagcgccg	cccgcgcg	27840
agcgaccggg	cgccgagccc	gccagcgacc	tgctcgtag	ctcgccgaag	acgcccgcg	27900
cactcgacgc	ggcggtcgcg	aacctcgcg	gtcatctcga	ggcgaccccg	gacgtcagca	27960
tcgcccagct	ggcgcacacg	ctgcgcgcgc	gccgacagga	gttcccgcat	cgacgagcgc	28020
tcgctgcac	cagccgcacc	gatgcgatcg	ccgtgttgcg	cggcgccgac	gcccacgtg	28080
tcgccaccgg	cgacgccgaa	tccggcagcc	gtgcccgtgg	cttcattgtc	tcggcgccg	28140
gcacgcagta	cgcgaacatg	ggccgggagt	tgctgcgggt	cgcgcagggt	ttccgcgaag	28200
aaaccgagcg	ctgcctcacg	caacctgtcac	tctccgcgca	cgtccgcgct	gcgtgtttc	28260
cggatcatgc	ggacgtcgag	cgggcgcgcg	attggctcga	agctccgtcc	gttggtttgc	28320
ccgcgctctt	tgctgtcgag	tacgcgctcg	cccgatcag	gatggcggtg	gggtgtgagc	28380
ccgacgggct	catcggtcac	agcatgggcg	agtacacggc	cgcgtgcctt	gcggcgctga	28440
tgctgctcga	gaacgcgctc	gcgctcgtgt	gtctgagagg	ccggctgttc	gagacgcttc	28500
cggccgggga	gatgctggcg	gtcccgtctc	ccgaggcaga	gggtcgcccc	gcgctcggcg	28560
atggactcgc	gatcgcccg	gtgaatgcgc	caagctcggt	cgtcgcttcc	ggcgccccc	28620
acgcgctcgc	ggcctttgcg	cgcgtgatgg	ccgaggccgg	tgctgaatgc	cgcgcctcc	28680
acatcgcggt	cgcggcgcac	tcgcccgatg	tcgatccgat	cctggcgagg	ttcgagcgtt	28740
tcctccgcgg	catcgcgctc	gacgcgcaaa	gggttccagt	cgtgtcgaac	ctcagcggt	28800
cgtggctgac	cgcggacgaa	gcgaggtcgc	cgcagtactg	gacgcggcat	ctgagacaaa	28860
ccgtcagggt	cttcgacggg	ctcggcacgc	tgctcgccga	gccgaatcgg	gtgctcctcg	28920
aagtcggggc	cggacagacg	ctgtccggat	tcgcacggca	gcacccacag	aaaacggcgg	28980
cgcacgtcgt	tcaagcgacg	ctccggcacg	ccaaggaacc	ggcgccggac	gtctgtttcc	29040
tgcagcaggc	gggtggcgag	ctctggacgg	cagggtgtcc	cgtgaactgg	gctgcttaca	29100
acagaggcga	gggcaggcgt	cgtgtgccgc	ttccggccta	tcggttcgag	aagaagcggt	29160
acctggtcga	atcgccccag	gcgcgggtcc	cagcgccgct	tcaagccctg	gcgccaaggg	29220
ccgatgctac	tcagcagatg	acgacaactg	tgcttctgct	tcgggcccga	tcgctgccc	29280
atcgatcccg	cgagcggtcg	accgaaatcc	tgcaacaact	gagcgccctc	ccggcagcgc	29340
agatcgatcc	cgcgctgtcg	tttctggaga	tggtattcga	ctcggtgttc	ctgacgcagg	29400
cgagctcgcc	gttcaaggcg	tggttcaagg	tcgggatcac	gttccgccc	ctcttcgaag	29460
aggcaccgac	gatcgacgcg	ctcacgagct	acatcgacgc	caagctgcct	gcggacgcgt	29520
ttcctgcgcc	ggcggaagcc	gctccggcgc	ctttcgtggc	cgcgaagcgt	ctgtcccgcg	29580
gccctcatgc	ggccctgggt	ggcgtgcccg	cggccgcgcc	cggcacgctg	cagcacgtga	29640
ttcaacagca	attgcagttg	atggccgagc	aactccgcat	gctgggcgga	aatccccccc	29700

tgctgcccgc	gatgccgccc	ccagcgctcgt	ccgtcgcggc	ggccgacgcg	ggactgaagt	29760
cccgcgctac	gacggcgccc	gtctcgaagc	cgcatgcggc	cttcaagccg	atcgacggca	29820
aagccagtg	ggagctgtcg	cccgcgacga	tgccgcgctg	cgccgcctgg	atcgatcgat	29880
acacacggcg	cacggccgga	tcgaagcgac	gcttggccga	gtaccggccg	gtgcttgccg	29940
atccacgctc	ggtcgcccga	ttcaaccggc	tctggaagga	gatgatctac	ccgatcatca	30000
ccgaccggtc	ggaaggatcg	aaggtctggg	acgtcgacgg	caacgagtag	gtggatctcg	30060
tgggcccgggt	cggtctcgctt	ctgttcggac	accgttcgcc	gatcgtagtc	gacgcccgtg	30120
aggaacaact	ccaccggggg	ttcgagctcg	gaccgttgcc	gccgctcgcc	ggcgaggtcg	30180
cggcgctcgt	caaggaattc	acggggatgg	agcgcggttg	attctccaac	accggctccg	30240
aagcgtgct	ggcgccacg	cgttttgccc	gaacggtcac	cgggcgcgac	aagatcgccg	30300
tcttcgaagg	gtcgtatcac	ggtctccttg	acgaagtgt	cagccggccg	ctcgtggtga	30360
acggcgaatt	gcggtcgccc	cccgcggcgc	ccggcatcgc	cggcagcgcg	gtgagcgaag	30420
tcatcctgct	cgagtacggc	aacccgaagt	cgctcgaggt	catccgcgcg	cacgagtcgg	30480
agatcgccgc	gatcctcgtc	gagcccgtgc	agagccggcg	gctcgacctg	cagccgaagg	30540
agttcctgca	cgaggttcgc	cgctggcgcg	acgagatcgg	cgccgcgctg	atcttcgacg	30600
aagtcacac	cggttcggcg	gtgcatcccg	gcgccgccc	ggcgacttcc	ggcgctcccg	30660
cggatcttgc	cgccacggcg	aaggtcgctg	gcgccggcat	gcccacggcg	atcgctcgcg	30720
gcaatcgcaa	gttcatggac	gcgctcgacg	gcgccgactg	ggagtagcgc	gacgggtcgt	30780
tcccggaggt	gggggtcacg	tactttgccc	gcacgttcac	gcgccatccg	ctcgcgctcg	30840
ccgctcgcaa	ggccgtgctg	acgtacatga	agtcgcaggg	tcccggcctg	cagacgcggc	30900
tgccagatcg	gatcgagcgg	ctcgcgaaac	acgcgcgcgc	gatcgtagcc	gaacttcccg	30960
cgccgatca	gattacgcag	ttcagctcga	tgatgagcct	gaacttcccg	cacgaccaga	31020
agacggcggt	gctgctgttc	ttcctgatgc	gcgagcgcg	gattcatatc	tgggaggggc	31080
gtcccttctt	cttcacgacg	gtcacacccg	aggccgacta	cgccgccatc	ctccgcgcgc	31140
tcgacgagtc	gctggccgag	atgcaggcgg	ctggcttctt	cgccgcgcgc	gcccactccg	31200
cattgcaggt	gacggggacg	gcggaaccag	acgtcctgcc	gttcaccgaa	ggacagcagg	31260
agatctggct	cgccggccag	atggggcgag	cgccatcacg	cgccataaac	gaagtcgtcg	31320
cgctcgacct	gcgccggcct	ctgaatcgcg	cgccgatgca	gcccgcgcgc	gacgaggtcg	31380
tcgcccgtca	cgatcgcgtg	cgccatgacg	tgccagcgga	accgctcgcc	ctgctgcacg	31440
ttccgggcac	gtccgtgccc	gttgaatggc	aggacgtgtc	gggattgccc	gaggacgctc	31500
agcacgaggg	cgtgcagaag	gtgctcgagg	gcggcgacgg	cgtggcgctt	gacgtggaac	31560
gcggggccgt	gctgcgcgtg	accaccatga	cgctcaccgc	cgagcatcac	gtcttcgtca	31620
tgccggccca	ccatctcgtg	tgtagcggct	ggtcgttcgg	cgctaccccg	cgcgatctgg	31680
cgacgttcta	ttcggcccag	gttcgaggca	cgctgtagaa	cctcgaggcg	ccgatgcagg	31740
tgagccgctt	cgccgctgag	gatcgcgagg	ccaagcagag	tcctgaagcg	gcccagaccg	31800
aggcggttct	gatccggcaa	ttcgacaccg	tcccgcgagc	gctcgagtta	cccgcgcgac	31860
ggccgcggcc	gccagtgaag	tcgtaccagg	gcccgcgtgt	gtccgtgccc	tttgacgcgg	31920
cgctcgctcg	cgccgccaag	aagctcgccg	cggaacaccg	cacgaccctg	tttacgacgc	31980
tgctcgccgt	cttcgagacg	ctcgtctacc	gactgaccgg	ccagcaggat	ttcgtcgtcg	32040
gcatcccggt	cgccgtgcag	cctctgctcg	gagaggtat	ggtcgcgcac	tgctcaact	32100
tcgtgcccgt	tcgcccgcgc	gtgtcaccga	ccgcgacgtt	ctcgagtagc	ctgaacacgc	32160
tgaggacgca	gtcactcgac	ctcaatgagc	accagaactt	cacgtacggc	agtttgctcg	32220
ccaagctgcg	gttgccgaag	caccggggcc	gagatccgct	cgctcgcagc	agcttcacgc	32280
tcgagcccgc	catcgtcgat	cccgggttcg	acgggtctac	ggcgcgcttc	ctgacggttc	32340
cgccgctgac	atcgaagcgc	gacctgcacg	tgaacgtcat	ggagatggac	ggcggtctgc	32400
tggtcgaggg	cgactacagc	accgatctgt	tcgacgaacc	gacgggtgcg	cggtggctgg	32460
atagctatcg	gatcctgctc	gagggcgtcg	tgctcgtccc	cgccgggtcc	ctcctgagcg	32520
tgccgggtgat	ctcggagccg	gaccggagtc	agcttgctac	cggtcggaac	gacacggccg	32580
ccgactaccc	gcgcgatcgc	gtcatccatc	aactcgttga	agagcgcgcc	gtgcagaccc	32640
cggagggact	cgccgtcgtc	tgtggtgccc	agcggtgac	ctggaccgag	ctcaaccgac	32700
gcgcaaacgg	gttgccgcat	gcgctcgtga	agaagggcgt	cgcccccggg	agcagggtgg	32760
cgctgtgcct	ggaacgatcg	gcccgggtta	tcgtctcggt	gcttgccgtg	cacaaggccg	32820
gcgcgcgata	tgtaccactc	gaccccgctc	cgcccgccga	tcgcaagtcc	ttcgtcgtcg	32880
aggattccgg	tgccgtgctg	gtcgtgacgg	attcgcggtc	cgaaggccga	tggtgacgc	32940
ccaggacgcc	cgtgctccac	ctcgacggcg	acagtccgcg	aatcgcgcaa	gagtcgcacg	33000
atgatctgaa	ggtcgccctg	tcggccgaag	acgcccgcgt	cgtgatctac	acgtccggat	33060
cgacgggaca	accgaagggc	gtcgtcatcg	ggcaccgcca	gctcgtgaat	tacacgtggg	33120
cggtgatcga	acggatcggg	ttaccggccg	gctcgaagga	tgccgtcgtg	tcgaacgtcg	33180
cggccgtag	cggtgtgacc	gggctctcgg	cgctcgcgat	gggctgggtg	ctgcacgtga	33240
tcacctcaga	ggctcgcgacc	gacggggcgg	cgctcgggtc	gtatttgacc	gttcacgaca	33300
tcgacttctt	gaagatcacg	ccgtcgtcgc	tcgctcgtcg	gctcggcgat	caccgctcgg	33360
cctcgtcgtg	gcccgcgcgc	tgtctcatgt	tggggcgcca	gccgtcacgt	ccggcggtgg	33420
tggaagagct	gcagcgtctc	gcaccaggct	gcggatcat	gaaccactac	ggcccgacgg	33480
agaccacggg	cgccgtgttg	acgtactgga	ccgggtgacc	cagggatctg	ttcgtcgaga	33540
gcgtgcccgc	cgccagggcg	ctcgggaacg	tccgcgccta	cgtgctcgat	ccggccggag	33600
cgcccgcgcc	gattggcgct	gtcggcgagc	tctgtatcgg	cgccgcgctg	gtggcacgcg	33660
ggtatctcaa	tcgtcccagc	ctcaccgccc	agcgcttcgt	ccccgatccg	ttcgccaccg	33720
agccgggagc	gcgcgatgtac	cggaccggcg	acccgcccgc	gttcggggcc	gacgggaaca	33780
tcgagttcct	cggctcgcgcg	gactatcagg	taaaagtacg	cggattccgc	gtcgagctcg	33840

gcgaaatcga	agcggcgctt	cgcgctcatg	agggcgctcga	gcagggcgta	gtcgtgctgc	33900
gaaaagatca	gcccggcgac	gagcgccctga	tcggtctatgt	cacgaccggc	ggcgggcgct	33960
cgatctcgat	ggcggagctt	cgcacgtcgc	tcaagcagag	gttgccggac	tacatgggtc	34020
cggcggtgat	cgtcgcgctc	gacaagctgc	cgggtcacgtc	gattggcaag	atcgatcgac	34080
gcgccctgcc	ggcgccgccc	gagcgcgcca	gtttcgagtc	ggagttcgtc	acgcccgcga	34140
ccgagacgga	gcgcgggctt	gcggagatct	ggagcgcggt	gctcgggtctc	gagaagattg	34200
gcgcgcacga	caacttcttc	gacctgggcg	ggcactcgct	gctcgcggcc	cgcaccgcca	34260
tgaaggtccg	cgacacgttt	cgcgtggacg	tgctcgtcat	cacgttcttc	gaaatgcccc	34320
ccatcgccat	gatggcgcca	gcaatcgacg	ccggtaccgc	ggcccccgcc	tccccggacca	34380
tcacgcccag	ggcccccgctc	agggccccgca	gggatgagct	ggtccgcccc	tgaacccccg	34440
ccaccccgag	actgagctcc	ccgacctgcg	gcgcgacggg	atcgccgccc	aagcgggcga	34500
cgtgtttctg	gcgcggcgct	cgtacgcccc	ggggcgcttg	tggttctctg	cgcagctcga	34560
gcctgaaagc	tccgcctaca	acattccgat	cgcgctcaga	ctgaaggggc	gcctcgaccc	34620
cgcccgctc	gaacgcgcca	tcaacacat	cgtccagcgt	cacgagacgc	tgccggacgac	34680
gttctcgctg	cagggcgatg	agctcaagca	ggtcatctct	ccgacgctgc	gtattccgat	34740
ctccctcgtg	gacatgcgtc	tgctcagctc	tgagaaacgc	gaggccggat	ggcgcccgct	34800
tgccggccgac	gacggcgcg	gcgtcttcga	tctggctgag	ggccccgctgc	tcaaggtcgt	34860
actcgtcatc	ctcgagccct	ccgatttctg	cctcctgac	accgtgcatc	acatcgtctt	34920
cgacgggtgg	tccggcgcca	tcttcatccg	cgagctggcc	gaggcgctacg	cggctgagct	34980
cgacggggcg	gacccgaagc	tgccggagct	gccgatccag	tacgcggaact	ttgccgcgtt	35040
ccagaaggaa	cagaccgaag	ggccccgcct	gccggtgctg	ttgaagcaact	ggacgcagga	35100
gctggcgggg	gcgtcgatga	cgctcgacct	gccgacggat	cgtccgcgtc	cgcccccgca	35160
gacctcccg	ggcgccctcg	cctcgttctg	gctcgacgga	gcgatctggc	cgggcgctgc	35220
cgcattcagc	cggggcgaga	acgcgacggt	gttcgatgacg	cttctctccg	ccttcggcgc	35280
cgtgctgcat	cgctgggtccg	gacaggaaga	catcctcatc	ggcagcccg	cggccggccg	35340
cgatcgtccg	gagcttgagg	gcctgatcgg	gttcttcatc	aacacgttcg	tcctgcggct	35400
caacgcgtcg	ggagatccca	cgtttcgcca	gttcgctcgg	cgggcgcgcc	gcgcgtgccc	35460
tgccggcctac	gcgcacacag	acctgccgtt	cgatcggtc	gtcgagacgc	tcaacccgga	35520
gcggaccggg	gctgcgcatc	cgatctatca	agtcagtgtc	gcgcaggatc	cgccggcgct	35580
gaagatgacc	ttcgccggga	tcgagctcga	gcggcttgct	gtccacaacc	gcagcaccaa	35640
gtgcgacctc	acgctcgagg	ttgccgaaga	cgccgacggc	gtcaccctgt	acgtcgagta	35700
cagcgtcgac	ctgttcgatg	cggcgacggg	ggcgctcgct	ctgcagcaat	atcgccgtgt	35760
gcttcagacc	gcgatcgcg	acccggacca	gcggctctcc	gagctgcgtc	tgcttgccga	35820
cgacagcg	atgcagctga	tcgcactggg	gaccggtccg	gcggccgtct	acccgcgcga	35880
cgtcagctcg	gccacggtct	tcgaagagcg	ggtccgcgcg	actccggcgg	cagtgggcgc	35940
gacgctggag	ggacagcacc	tcacgtacgc	ggatctcaac	cggcaggcga	accgcctggc	36000
ccggcggttg	aaggcggttc	gtgtggggccc	gaacgtgctg	gtcggcggtg	cgctcgaccg	36060
gtcgttcgat	ctgctggtgg	cgctcctcgc	cgtggcggaag	gcccggcgcg	cgtatctccc	36120
gctcgatttc	gctcaccgc	aggaacggct	tgcgctcatg	ctcgccgaca	cgcgcgcgcc	36180
catcgtgctc	acgctgcggc	gatttgccgg	ggcgctcgag	tcctttcccc	tcttctgcct	36240
ggacgacgag	ctgccggcca	ccgccgagga	gccggacgaa	gatctgccgc	cgagtcgac	36300
cggagaccac	ctggcctatg	tgatgtacac	ctcgggctcg	actggctcagc	ccaagggcgt	36360
cgcgatcacg	catcgcgcg	tggttcgtct	cgtacgcggc	acagactacg	tccgctggaa	36420
cgacgtgcgg	gcgatcctcc	acatcgcgcc	gacgtcgttc	gacgcgttga	cgttcgaggt	36480
gtggggcgcg	ttgctgaacg	gcgcgcggct	ggtgctggtt	cccagcgggc	tcgtcagcct	36540
ggagacgctc	gagcgcacgc	tccggtcgga	gcaggtggac	tgcttgcgtc	tcacgaccgc	36600
gctcttcaac	gccgtcatcg	acgcgaagcc	cggcattctc	gtcggcggtga	agcaactcct	36660
gacgcggcgc	gaagcgctgt	cagtcgcgca	cgtgcggcgc	gccctggcgc	acctgcggga	36720
cacgcggctc	gtgaacgcgt	acggtccgac	cgagtgcacg	accatcgcg	gcgcctacca	36780
gattccgcgc	acgctcgatc	cgcaggcgcg	gtccatttct	atcgccgggc	cgattgcca	36840
tacgcaggcg	ctcattctcg	atcgacacgg	cgacctcggt	ccgttcggca	tcgccggcga	36900
gctctgcctc	ggcggagacg	ggctggcgcg	ggaatacctg	aatcagccgg	cgctcactgc	36960
ggaacggttc	gtgcccaacg	cgttcagccc	cgagccggcc	gcccggctgt	accgcacggg	37020
agacctcgcg	cggatgcgcc	gcgacggcaa	catcgagttc	cttggccgga	tgatcgccca	37080
gatcaaaactg	cgcggcttcc	ggatcgagcc	gggcgagatc	gagcatgccc	ttcgtgcccc	37140
gggagacggt	ctggacgcgg	tggtggagat	ccgggtcgat	gccacatgg	cccatctcgt	37200
cgcctacgct	gtgcgcggcg	atgggagtc	actcaccgga	accgacctgc	gcgaacgtct	37260
gaaggcgcg	ctgccggaat	actgtgtgcc	gcggaagtac	gtgttctctg	ccgaggtccc	37320
gcgaacggcg	gcccgcgaag	tcgatcgctc	ggcgctcct	gcacgcggcg	tcgagggccc	37380
cgcggccgag	cgcgcgcatc	cgctgcttct	gaacgaggtc	gaagcggagc	tcgcgcgaat	37440
ctggagcgaa	gtgcttcggc	tgccctccgt	gctcgcgacc	gacaattttt	tcgatcgtgg	37500

21/39

Figure 5b

```

CDS      complement(3..914)
         /blast_score=2e-66
         /blastp_match="AE004644.PA2177 Pseudomonas aeruginosa"
         /gene="regulator hybrid"
         /note="probable similarity to prokaryote sensory
transduction proteins"
         /product="sensor/response regulator hybrid"

CDS      924..2168
         /note="none"
         /gene="ligase"
         /blastp_match="AP003013.MLR8297 Mesorhizobium loti"
         /blast_score=e-152
         /product="2-amino-3-ketobutyrate CoA ligase"

CDS      2207..3190
         /blast_score=e-151
         /blastp_match="AE008872.TDH Salmonella typhimurium"
         /note="none"
         /gene="dehydrogenase"
         /product="threonine 3-dehydrogenase"
         /pfam_match="PF00107; adh_zinc; 1"

CDS      3373..4455
         /note="putative"
         /gene="methyltransferase"
         /blastp_match="AE001866.DR0026 Deinococcus radiodurans"
         /blast_score=1e-08

CDS      4546..4959
         /blast_score=2e-11
         /blastp_match="AP002997.MLL1617 Mesorhizobium loti"
         /gene="unknown"
         /note="pfam00263, GSPII_III, Bacterial type II and III
secretion system protein, Expect = 7.8"

CDS      5176..6192
         /blast_score=5e-98
         /blastp_match="AF064070.PE20 Burkholderia pseudomallei"
         /gene="glucose epimerase"
         /note="putative "
         /product="UDP-glucose 4-epimerase"
         /pfam_match="PF01370; Epimerase; 1"

CDS      6331..14043
         /note="substrat AT, malonyl ; zinc depend dehydrogenase ;
zinc dependent adenosine deaminase putative"
         /gene="PKS I"
         /blastp_match="AF285636.WCBR 2547 Burkholderia mallei"
         /blast_score=0.0
         /pfam_match="ketoacyl-synt"
         /pfam_match="ketoacyl-synt_C"
         /pfam_match="Acyl_transf"
         /pfam_match="SAM binding"
         /pfam_match="adh_zinc"
         /pfam_match="pp-binding"

CDS      14275..15408
         /blast_score=e-104
         /blastp_match="AF285636.WCBT Burkholderia mallei"
         /gene="acyl-CoA transferase WcbT"

```

22/39

```

/feature="CDS"
/feature_start="15436"
/feature_end="16245"
/feature_qualifiers="
  /note="putative"
  /pfam_match="PF00155; aminotran_1_2; 1"
  /blast_score="5e-11"
  /blastp_match="TTDEFFMT.FMT T.thermophilus"
  /gene="formyltransferase"
  /note="evidence experimental"
  /product="methionyl-tRNA formyltransferase"
  /pfam_match="PF00551; formyl_transf; 1"
  /pfam_match="PF02911; formyl_transf_C; 1"
  /note="putative"
  /gene="glycotransferase"
  /blastp_match="AF285636.WCBD Burkholderia mallei"
  /blast_score="e-99"
  /blast_score="2e-82"
  /blastp_match="AF285636.WZM Burkholderia mallei"
  /gene="ABC-2 transporter Wzm"
  /note="putative"
  /pfam_match="PF01061; ABC2_membrane; 1"
  /blast_score="7e-61"
  /blastp_match="AF285636.WZT Burkholderia mallei"
  /gene="ABC-2 transporter Wzt"
  /note="putative"
  /pfam_match="PF00005; ABC_tran; 1"
  /blast_score="1e-26"
  /blastp_match="AE009248.ATU3189 Agrobacterium tumefaciens"
  complement(21164..24301)
  /note="transporter domain"
  /gene="unknow"
  /blast_score="5e-29"
  /blastp_match="AE009122.ATU1658 Agrobacterium tumefaciens"
  /prosite_match="PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1"
  complement(24351..27023)
  /note="none"
  /gene="unknow"
  /blast_score="5e-29"
  /blastp_match="AP003581.ALR0267 Nostoc sp"
  complement(27806..29686)
  /note="none"
  /gene="cell surface protein"
  /blast_score="2e-34"
  /blastp_match="AE010748.MA0851 2567 Methanosarcina
  acetivorans"
  /product="cell surface protein"
  complement(29535..30872)

```

23/39

CDS /note="none"  
/gene="cell surface protein"  
/blast\_score=2e-36  
/blastp\_match="AE010748.MA0851 Methanosarcina acetivorans"  
complement(30848..32647)  
/note="fragment"  
/gene="O-antigen"  
/blast\_score=1e-72  
/blastp\_match="AF105060.RFBC Riftia pachyptila"  
/product="O-antigen biosynthesis protein"  
/pfam\_match="PF00535; Glycos\_transf\_2"  
complement(32574..35555)  
/note="putative"  
/gene="glycosyltransferase"  
/blast\_score=7e-46  
/blastp\_match="AE013462.MM2213 Methanosarcina mazei "  
complement(35533..36598)  
CDS /blast\_score=7e-37  
/blastp\_match="AE013462.MM2213 Methanosarcina mazei"  
/gene="glycosyltransferase"  
/note="putative"  
complement(36516..37400)  
CDS /blast\_score=8e-22  
/blastp\_match="AP003581.ALR0267 Nostoc sp"  
/gene="ALR0267"  
/note="putative ATP/GTP-binding protein, esterase fush  
9e-14"  
Sequence 37507 BP; 5531 A; 13507 C; 13011 G; 5458 T; 0 other;

gatccacga	tcggcgccag	cggattgcgc	agctcgtgcg	cgagcatcgc	gaggaactcg	60
gtcacgtggc	ggccctcggc	ctccagctcc	tccatgcgct	tgcggtgcga	gaggtcgcgc	120
gtcaccttgg	cgaagcccat	gtgccttcca	ctcgcgtcgc	gaagcgccgt	cacgacgacg	180
tcggcccaga	agcgcgtgcc	gtccttgcgc	acgcgccagc	ccgtgtcctc	ggcgcggcgg	240
tgcagcagcg	catcgtgcag	ctcggcctcc	gggcgccggg	ccgcggcatc	ctcggggcga	300
tagaagagcg	aaaagtgcgc	tccgatgatc	tcttcggcgc	tgtagccggt	gatgcgctcg	360
gcgcccctgt	tccagctggg	cacgtggccc	gcgggggtcca	gcatgtagat	cgcgtagtcc	420
tgcacgccct	cgatcaggag	gcgcgagcgc	tcctcgcttt	gccgcagctg	ctcctcgtgc	480
tcgcggcgct	cggtcagggtc	gcgcgtgacc	ttgccgaagc	cgaccagcgt	gccgtcctcg	540
tcgcgcagcg	ccgtgatgat	gacgttggcc	cagaagcgcc	ggcgtcctt	gcgcacgcgc	600
cagccctcgt	cctcgaagcg	gccgtcctcg	cgtgcgcgca	gcagctcctt	ggcgggccag	660
tcgcgcgcga	tgctcctctc	ggtgtagaac	gcggagaaat	ggcgccgat	gatctcttcg	720
gcgcgatagc	ccttcaactg	ttcggcgccc	gcgttccacg	aacgcacgta	gccctcggga	780
tcgagcacga	agatcgcgta	gtccgtgacg	gacccgacca	gcagccggaa	gatggcgtcc	840
tcgccggcca	gcccgtcgcg	ggcgcggaag	gcgaggttgc	gcaccgaggg	ccgcggaccg	900
gatgggcttg	gcatggctat	tagggtagaa	cggtgcacgc	aagttgcaat	cggacaatca	960
gagataattg	cttgatgaag	atgccatttt	tggaattctt	acgcaggag	ctggggacgc	1020
tcgaggccca	gggcctgtac	aagtccgagc	gcgtgctcga	ctctccgcag	ggcgccagcg	1080
tgcaggtggc	cgcccacgac	gtgctcaact	tctgcgcca	taactacctc	ggcctggcca	1140
accaccggc	cctggtcgcg	gccgcgaagg	agacgctcga	gcgcgacggc	tacggcatgg	1200
cgtcggttcg	cttcatctgc	gggacgcacg	ccgtgcaccg	cgagctggag	acgcgcctcg	1260
cgcgcttct	cggcaccgag	gacgcgatcc	tgtagcgctc	ctgcttcgac	gccaacggcg	1320
ggctcttcga	gacgctgctg	ggcgaggagg	acgcggtgat	ctcggacgcg	ctgaaccacg	1380
cctcgatcat	cgacggcggtg	cgcctctcga	aggcgaagcg	cttcgcgtac	gacaacaacg	1440
acatggcgag	cctcgaggcc	cagctgaagg	cggcggaagc	ggccggcgcg	cgctacaaga	1500
tgatcgccac	cgacggcggtg	ttctcgatgg	acggcatcgt	cgcgaaacctg	ggcgccatct	1560
gcgagctggc	cgggcgctac	cacgcgatgg	tgatggtgga	cgactcgcac	gccgtgggct	1620

24/39

tcattgggccc	gcacggggcgg	ggaacgcccg	agcattgccc	tgtcgaaggg	aaggtggaca	1680
tcctcaccgg	cacgctcggc	aaggccctgg	gcggcgcttc	cgccggctac	acggcgggca	1740
agcgcgaggt	ggtggcctgg	cttcgcaacc	gctcgcgccc	ctatctcttc	tccaacacgc	1800
tgatgcccgc	catcgcgggc	gcgtcgctca	aggtgctcga	tctcctcgag	ggcggcggcg	1860
agctgcgcgc	gaagctcgcg	cgcaacgccc	gccacttccg	cgccgagatg	acgcgcctcg	1920
gcttcacgct	ggccggcgcc	gaccatccga	tcattcccgt	gatgctgggc	gaggccccgc	1980
tcgcgaagga	gatggcggac	cggctgctga	aggaaggcat	ctacgtgggc	ggcttctcgt	2040
ttcccgtgg	gccccgaggg	caggcgcgca	tccgcacgca	gatgtcggcc	gcccataaac	2100
cgaagcacgt	cgatcgcgcc	atcgccgcct	ttgccaaagt	cgggcgcgat	ttgggagtc	2160
ttgcttgaga	accctttcga	agacgaagcg	cgagcccgcc	atctggatgg	tcgagtcgcc	2220
caagcccggt	gtcggccaca	acgacgtgct	gatccgcgtg	aagaagaccg	ccatctgcgg	2280
caccgacatg	cacatcttca	actgggacga	ctggctcgag	aagaccatcc	cggtgccgat	2340
gacggctcgg	cacgagtagc	taggcatggt	ggaggccatg	ggccaggagg	tgcgcggcct	2400
gcaggctcgg	cagcgctctt	ccggcgaggg	ccacatcgct	tgcggccatt	gccgcaactg	2460
ccgcgcgggg	cgccgcccac	tgtgccgcaa	cacgcagggc	gtggggcgta	accgccccgg	2520
cgcgcttcgg	gactacctgg	tgattcccgc	ggagaacgcg	ttcccgatcc	ccgacgacat	2580
ccccgacgag	atcgccctga	tcctcgatcc	gttcggcaac	gcggcgacac	ccgcgctctc	2640
gttcgacctg	gtgggcgagg	acgtgctgat	caccggcgcc	gggcccgatcg	gcacatgggc	2700
cgcgccatcg	gcgcgccacg	tcggcgccgc	ccacgtcggt	atcacggaca	tgaacgacta	2760
ccgcctggcg	ctcgccacga	agatgggtgc	cagccgcgcg	gtgaacgtat	cgaaggaaaa	2820
cctgaaggac	gtgatgcgcg	agctcggcat	ggtcgagggc	ttcgacgtcg	gcattggagat	2880
gtcgggctgt	ccctcggcct	tccgccagat	gctcgacacc	atgaaccacg	gcggcaagat	2940
cgccatgctc	ggcatcccgc	ccagcgaggg	cgcgatcgac	tggaccagg	tgatcttcaa	3000
gggcctgggt	atcaaggggc	tctacggggc	cgagatgttc	gagacctgg	acaagatgat	3060
cgccatgctg	cagagtggcc	tcgacctctc	gcccattggt	acgcaccgct	tcgacgtgctg	3120
cgagtacctg	aagggcttcg	agacgatggg	ctccgggaaa	tcggggaaa	tcgtcctgtc	3180
gtgggattag	cgaggggccac	ggatcgcggc	ccgggttccc	cgccggcgat	cggtcaggcg	3240
ggttggcagc	acgccccggg	atttccgcgc	tggtcacgt	gcgacttcgc	cgctcgacgca	3300
ccgatcgcca	tcgacgtcga	cgtgagcgtg	tcgctgctcg	atgcccagg	cacctcgctg	3360
tggacgcagt	cgatggccgg	cagcgggcgc	gacaccccgt	ggctaccgac	cggcacgtat	3420
cgcgctccc	tgcgcctgga	gccgttccgc	ttcccggccc	cgcccgccct	ggtcaggttc	3480
gcgctctcga	tgcgggtgca	aggtgagcgc	cgctcgtgg	ccaccgcagc	ggcggtgggtg	3540
cccgcggggc	cacccacagg	cgtggcgcg	cgccctggc	acctcgaggc	cctggatggc	3600
acgcggcgcc	tcgagacgct	ggcctggctc	gatcccacgc	acagctgggt	ctcccagcat	3660
ttcgaccatg	cgctcgcgac	gtgcgtcgag	tacctggggc	ccggatggcc	cggttgccgc	3720
ggcgcgctgc	tggatgcggg	atgcggcgac	ggcatcacgg	cgctcggcat	cgcgctgcgc	3780
tatgcgccc	agcaggtgg	gggcatcgat	ccggggcgct	gctaccgggt	gcttcccagc	3840
atcctcgagc	gccacgccc	cggccacctc	gcgctgcgg	acaacctgca	gttctctccc	3900
gcggatgcc	acgcgctccc	gttcgcccgc	agcagcttcg	acgtcatcgt	ctcgtggctc	3960
gcggtcgagc	atttcgtggg	cggctacctg	ccctcgctcg	ccgaggcgcg	gcgggtgctc	4020
aagcccgggc	gcctgctgg	catccatccc	gagctctact	acaccgcgca	ccacggccat	4080
cacctcggcg	agtagagccg	cgagcccttc	ttccacctgg	tgaagtgcgc	cgacgaggtg	4140
cgcgacatcg	tgctcggcgc	ggacgtgagc	ctccacgacc	gcggcgggat	cgcccgagcg	4200
cgcgcgagc	actggcgctg	gtacaacgag	ctcaatcgca	tcacggccgc	gcagttcgag	4260
gacgagttgc	gcgcgctcga	cttcgagccg	tggcgccctg	cactgcgcgc	cgagccgctg	4320
gtggagtaca	agccggagct	gctgcgctac	cgcttcacgg	agctgggctg	ctccgagctc	4380
tacgtggcct	gcatacaacc	caagcccgtg	tcagatgggt	cctacaggac	gcattgggag	4440
gcggcattcc	ggtaaggctc	gccgggacgg	gaattgaagg	cgcaagcgcg	ttgttatccc	4500
cgtaactccc	ggactgcggg	gatcaccttc	acaggagaac	ccccgatgaa	gacccagacc	4560
caggccgtgc	tcgcccgggc	cctgctgtcc	cttgccggca	ccacgctggc	ccgcgcccga	4620
gccgtcacgc	tgacggggcg	gagcgaagtc	ccgcccgtca	cctcgagcgc	gaagggcagc	4680
ggcacggtgg	tcgtgaaggc	cgattgcacg	gtcacccgca	agatcacctg	caccggcgatg	4740
accgccaccg	cggcccacat	ccacgaaggc	aaggccggcg	ccaacgcggg	cgtggcagtc	4800
ccgttcgtga	agaccgcgga	cacacggttc	gaggcagccc	cgggcgccaa	gatgaccgag	4860
gcgcagtgcg	ccgctataca	ggcggggcgc	acctacgtga	acgtgcacag	cgaggccac	4920
aagggcgcg	aggtccgcgc	ccagctggcc	ggcaagtagg	gcagccgctg	caaaaagaaa	4980
aagcccgcg	aagcgggctt	tttcgttgaa	ggctttggtc	ggggcgaaag	gattcgaacc	5040

25/39

ttcgaccccc	tgcaccccat	gcaggtgcgc	taccaggctg	cgctacgccc	cgaccgaact	5100
tgaattata	cctgcatcgg	ggtcgcagcc	gccatgaatc	gaaacgcccc	gggcttcggg	5160
tagcatccgc	ccatggcgac	gaccctcgtg	acgggtggtg	caggctacat	cggcacccac	5220
atcctctgcg	cactggccca	ggccgggcga	cgcagcatct	gcatcgacaa	ctactccaac	5280
agctcgcgcg	gttccatcga	acgcgtgctg	cagatcgcac	ccggttgctg	cgaggccttc	5340
gacgtggaca	tccgcgacgc	cgacggcatc	cgcaagggtga	tccgagggcg	cgacgtggac	5400
agcgtgatcc	acctcgcggg	actgaaggcg	gtcggcgcaat	cggtcagagca	gcccagagcg	5460
taccacgaca	acaacgtccg	cggcaccgag	agcctgctcg	cggcgctcgc	cgactcgccg	5520
gtgcgcaagt	tcgctcttcag	ctcctcggcc	accgtgtacg	gcctcgcgca	gaagatgccg	5580
atcgacgagg	acgcgcccac	ctcgcgcgag	agcccgtacg	gccagaacaa	gctcgacatc	5640
gagcacatgc	tggctgcacct	cgccaagcgc	gaccctcct	ggcgcgtagt	gaacctgcgc	5700
tacttcaacc	gggtcggcgc	acacgagagc	gcgctgatcg	gcgaggaccc	ggccgggggtt	5760
cccaacaacc	tcatgcccta	cgtgtgccag	gtcgcctccg	ggcgactgaa	ggagctgagc	5820
gtctacggca	gcgactaccc	caccccgac	ggcaccggca	tgcgcgattt	catccacgtc	5880
tgcgacctgg	ccgaaggcca	cgtggccgcg	ctggaagcac	tcgagcgcgc	ggcccccg	5940
accgtcctca	cgggtgaatct	cgggacgggg	ctcggctaca	ccgtgctgca	attgctcgaa	6000
accttcgagc	gcgtgaacaa	actaagcatt	gcccggcgca	tcggtggggcg	gcggccgggc	6060
gacgtcgcga	tctgctacgc	gaacgcgggc	gcgcggaagg	ccgtgctggg	ttggaccgcg	6120
aagcgcggaa	tgcaggagat	gtgccgcgat	gcctggcggtt	ggcaggaaag	gaatgcaaag	6180
ggctacgggtt	gacggattgc	aaatcatggc	cttggcgacc	cttgctgcaa	tgcggcatga	6240
agggggcggt	tcttgtaaga	tccaaggttc	gggatgcgtc	caggacgtca	aaaacagcac	6300
ccggcggttg	aaggaagaag	gcaaccctag	tgttgtcatg	gctccagacg	ggattgcgcg	6360
ccgagtaaga	atcgtcagtc	gagaccaact	cgcctccggg	tgaacgaat	cgtgagccga	6420
agcaaacact	cgcatacgct	cgcctcgtc	ggcattgcgt	tccgccttcc	cggcgctacg	6480
cgcgaatccc	tctggccctc	gctcctcgag	gggcgcaacc	tcgtcaccac	cgtcgatccc	6540
gcgcgctggg	cgaaggagcc	gtactaccac	ccgcgcaagt	ccgagcccg	caccagctac	6600
tcgttcgcgcg	ccggcaccat	cggcgacatc	agcggcttcg	aocgctcggt	cttcggcatc	6660
tccccgcgcg	aggccgcgca	gatggatccg	cagcagcgcc	tgtgtctcga	gctctcctgg	6720
gaagccctcg	aggccgcggc	ggtgaagccc	tgcaccctgc	gcggcagcga	ttgcggcggtg	6780
ttcgtggggc	tgtccaccgt	cgactacagc	tacggcctcg	cctacgacct	ggcgctccatg	6840
gacgcgtcgt	cggccaccgg	caacaccgcg	agcatcgccg	ccaaccgcct	ctcgatttcc	6900
tacgacctcc	acggccccag	catggctcgtg	gatacggcgt	gctcgtcctc	gctgggtggcg	6960
ttccaccagg	cctgccagtc	gatcgccatc	ggcgagacgc	gccaggcgct	ggtgggcggc	7020
atcagcctgc	acctgcatcc	gtacgggttc	atcgcgcttc	ccaaggcgct	gatgctctcg	7080
aagcgcggcg	cctgccgcgt	gttcgacgcc	agcggcgatg	gctacgtgcg	ctcggaaggc	7140
gcggggcggtg	tggtgctgaa	ggacctcgac	cgcgcgctgg	ccgacggcaa	cccgatcctc	7200
gccgtcgtcg	ccgcgagcgg	cgtgaatacc	gcggggccgt	cctccagcct	caccgtgcc	7260
agcgtcgatg	cccagggtgc	attgctgtcg	gatgtgtacg	agcgcgcggg	catcgacccg	7320
tccgagatcg	actacctcga	ggcccacggc	accggcaccg	tggtgggcga	tccgatcgag	7380
acccgcgcc	tcggcatcgc	gctcggggcg	ctgcgcccg	gccaccagcc	gctgccgatc	7440
ggttcggtga	agagcaacgt	cggccacctc	gaggcggcct	cgggcatggc	gggcctggtg	7500
aaggcgctgc	attgcctcaa	gcaccgcgcg	gtgccgccca	ccatccacct	cgattcgccc	7560
aaccccaaca	tccacttcga	cgagtggaa	ctgaagccgg	tcctcgagcc	gctcacgctc	7620
catccgtaca	agaagctcgt	cgtcggcggtg	aactcggttcg	gcttcggcg	cgccaacgcg	7680
cacgtgatcc	tgcagagcac	gcccacttcg	cccaaggcgg	tcgaggcccc	ggccggcgag	7740
gtggcgctcc	tgctgagcgg	ccacgacgac	gcggcgctgc	gcgacgtggc	ggcctcgtac	7800
gcgaacttcc	tcagggcgag	cgagcaaccc	gcgctgcacg	acatcgcgta	caccaccgtg	7860
cactgccgcg	actggcacca	gctccgcgcc	atgggtggtcg	gcaccgaccg	cgccagcatc	7920
gccggcgccc	tgcgcgaatt	cgccaagggc	cacaacacgc	ccggcgctcg	gagcggccgt	7980
gcgctgccca	acgcgtccgg	ccccgcattc	gtgtattcgg	gcaacggcgc	gcaatggccg	8040
gcaatggggc	gcgcctcgt	cgccggcgag	ccggtgttcc	gtgccgcgct	gcaatcggtc	8100
gacgcgctgt	tccgcccgt	cggcggttac	tccatcctcg	agatgctgga	cgaggaggtc	8160
cccgcgcgac	agctggaccg	caccgagatc	gcccagccgc	tgctcttcgc	cgtgcaggtg	8220
ggcttgaccg	agctgctcgc	ccactggggc	atccagccca	cgcgcgtcac	cgccacagc	8280
gtgggcgagg	tcgcgcggc	ttgggcgagc	ggcgcgctga	cgctcgacga	tgcggcgcg	8340
gtgatctacc	accgcagccg	ctaccagggc	gagaccaagg	gcgcggggcg	gatgaccgcc	8400
gtcggcctcg	ccgaggacgc	cgcgcgcgcg	cagatcgagg	cgctcgggtt	gaagggcgag	8460



26/39

gtgaacgtcg	cgtgcatcaa	cagcacgcgc	aacgtgacgc	tgcgcggcac	gcgcgcgcgc	8520
atcgaggcgc	tcgaggccga	gctcacgcag	cgcaaggtgt	tccaccgccc	cctcgacctc	8580
gactacgcgt	tccacagccc	ggcgatggac	ccggtgcgcg	acgggctcgt	gagcgcgctg	8640
cgcgccctca	cgccgcgcgc	caccgcgggtc	gcgttccatt	cggcggtcac	cggtctgccc	8700
gccgcgggca	accagctcga	cgccacctac	tggtaggagca	acatccgcga	gccggtgcgc	8760
ttccaggccg	cgatccgcgc	catcgccgag	tccggcggtca	atgtcttcat	cgagatcggc	8820
ccgcacccca	tcctcaagaa	ctacatcaac	gacggcctgc	gcgcgcgcctc	gatcgagggc	8880
cgcgcgctgg	tcacgctgca	gcgctcggcg	agcgatcgcg	cggcgatccg	cgccgcgcgc	8940
caggaagtgc	tgatcacccg	ctgcccgggtg	gagaccgcga	agatgctccc	cgcgacgggc	9000
cacttcgtcg	acgtgcccgc	ctaccctggg	cagcgcgagc	gccactggcg	cgcgcccacc	9060
tcgcaggcct	acgacctcat	ccagcacggc	aagcagcacc	cgctgctggg	ctatcgcttg	9120
cacgagaacg	acttccagtg	ggaaaaccac	atcgacaccg	cgctctatcc	ggcgtagcgc	9180
gaccatgtcg	tcggcgggcg	cgtgggtgttc	cccgcggccg	gcttcgtcga	gatggcgctc	9240
gccgcctcgg	cgatcgcgct	gggcggcgac	gcgcacgaga	tcgagacgct	cgagatccgc	9300
agcgcgctgc	tgctcgagga	ctccacgtcg	aagaccgtgc	gcttcgcgct	ggaacccgat	9360
ggccgcttca	caatccgcag	ccgcgctcgc	ctgagcgagg	atccgtggca	actccacgtg	9420
gtcggcaagc	tggtcggcac	gcccaccgag	cttcgcgcgc	cgccgcgcca	gtcgcgcgcg	9480
tcgcgcgaag	ccgatgtgat	cgcgggcccg	cactacgaag	gcgcgcgcaa	ggccggcctc	9540
gcctacggcc	cggccttcca	gtcgggtctcc	aaggtgtggc	tcgacgcccgc	ggacccgggc	9600
agcgccacg	cgcgctgat	cctccccaag	ccgattcgcg	gcgagctggg	cgtgatgcac	9660
ctgcaccccg	cgtcgctgga	cggtctgttc	cagctgctgg	tcgacctgct	gcgcgcccag	9720
gcctcccggc	acgcgcagat	cgcttctgtg	ccgatcctgg	tcgggcgcac	gcgcctctac	9780
ggccctgccc	ccctcgtcac	cgctgccgc	gtccgcctga	ccgcgcgcag	cccgcgctcg	9840
ctggtcgcgc	acttcgagct	ctacggcccc	aacggcgacg	tggtggccgc	gctcacccgc	9900
gtgcgcttcc	gcgctgtcca	gctgaagcac	ccgcaggcg	cgcgctgctg	ctacctgtcg	9960
taccgcgcca	tcccgcgcct	gaatggcggc	gacacgcggg	gcgccaagcc	gctgcccctc	10020
gcgttgctgg	ccgcggcggtg	cggcgagcgc	tggcattcgc	tctcggcccg	ctcgcgcaag	10080
cgctactacg	atgaggctga	gccgctctcc	gacgtgctgt	gcagctcctt	cgccgagcac	10140
gcgctgcgaa	agcggatcgg	cgaggccccg	ctcatcgacc	cgccgcgctg	cgccgcgctg	10200
acgcgcggccg	agcattggcc	gctcctcaag	cgcgcgatcc	agatgctggt	cgaggaccag	10260
ttgctcgagc	ccgccgaagg	cggttgccgc	tggagcccg	tcaccgaatt	gcccgcgcgc	10320
caggagagct	ggatcacgct	gctgcgcgac	taccggagc	aggccggcca	gctgatcgcg	10380
ctcgggcgcg	ccggcgctca	cctcgcgga	gtgttcgccc	gcaaggaaagc	gtccttgccc	10440
agcctgttgc	cgggccacgca	ggcgaaacgtg	ttcgcgcgcc	tgtgctcggg	cgaccccgcg	10500
ttcgccaacg	ccgggcgcgc	catcgccgac	acgctggcgc	tcgccctcga	gcgcctgccc	10560
gcgcaccggc	cgctgcgcgt	gctcgagggtg	acgccgtgcc	gctccgagat	cacgctgcgc	10620
gtcctcgagg	ccgtggacct	cgaccgctgc	gagtaccacg	tcggctgcac	cacggacgaa	10680
gactggggcg	agtacgaagg	cgtcctcgat	gccattccgc	acgtcgagac	ttgcgtcgtg	10740
gacctgaaga	tcccgggcat	gggcctgaag	gcgccgaacc	agggcccgtt	cgatgtcgtg	10800
atcgtttccg	atggctgtg	cggtgcggcc	gatccggatg	cgcccgcggc	gcacctcgcc	10860
ggcgcgctcg	cggaagacgg	gctgctgggtc	atggtcgcgc	agcatccgtc	gcgctggacc	10920
gacctcctct	tcggcgcgga	cccggcggtg	tggacgctcg	ggccctcggg	ctcgcagcgc	10980
tcgcgcctgc	gttcgcccga	cgagtggcgc	atcgcgctgg	cccacctcgg	cttcgggccc	11040
accgtcgcgg	tgcccagact	gcccggcctc	gagcgcggt	cgtagctcct	catggcgcg	11100
cgcgacggcc	cggtgccgca	ggctgccgag	cgcagcgttc	ccgccaagtc	gggctggatc	11160
ttcgtgcagg	aagcgggcgg	ttattcggcg	gcgttcggcg	cctgcgtgca	ggcgatctg	11220
cgcgacgcg	gccacgcggt	gttcgagatg	gcgcccgaatt	ccgtcgagg	ggccgagatg	11280
ggcgagcttg	cgcgcgcgcg	ccagacgctg	ggcgacatca	gcggcatcgc	cttcctggcc	11340
ggcctgcccc	acgacgcggc	cgccacgggc	gacgcggccg	agcgctcgc	gcgccagacc	11400
gcgcgctgcg	ccgcgctcgc	gcgcctgctg	cgcgaaatgca	agggagtg	catgacaccc	11460
gcggtgtgga	tcgtgacctc	gggcgcggcc	tccagcctgg	cgcgcaacgt	ccccgcctt	11520
gcctcgcgcc	gcccgcctcc	cgatccggac	gaggcgatcg	tatgggggtt	cgcgcgacc	11580
gcgatgaacg	aattccccga	gctgcgcctg	cgctcgtgg	actgcccga	cccggagaac	11640
ctcgagcgca	acgccgcctc	gctggtgcag	gagctgctct	cgcccaccgc	cgaccagag	11700
gtggtcctca	ccggcgccgg	ccgctacgcg	ctgcgcctg	gcgcggtgcc	gccgcccga	11760
tcggtgcgca	ccgctaccca	gggcatcacg	cgctcgaact	tcgccgcgc	cgcccgctg	11820
aagaacctcg	cgtggcgcg	cgacgtgcgc	cgcaagccgc	gcgcgaacga	ggtggaaatc	11880

27/39

gaggtgcgcg	ccgcgggcct	caacttccgc	gacgtgatgt	acgcgatggg	cctgctctcc	11940
gacgagggcg	tcgagggcg	cttcgcccgc	gcctcgctcg	gcatggagct	ctcgggcgtc	12000
gtcaccgcgc	tcggcaagga	cgtgacctcg	gtcgcgcgcg	gcgacgaggt	gctcgcgttc	12060
gcgcctcgcg	cgttcgccac	gcacgtggtc	accaccgcgc	attcggctcg	gaagaagccc	12120
gccggctgga	cgttcgagtc	cgcgccacc	atcccgaccg	cgttcttcac	cgtgtactac	12180
tcgttgaagc	acctggccca	gctgcgcgag	ggcgagcgcg	tcctgatcca	cggcgcgcgc	12240
ggcggcgtgg	gcatcgccgc	gatccaggtc	gccaaagtgg	tgggcgcgga	gatcttcgcc	12300
accgcgggct	ccgacgagaa	gcgcgaactt	gtgcgcctgc	tgggcgcgga	ccacgtgctc	12360
gactcgcgca	cgctcacgtt	cgccgacgac	gtgctgcgca	tcaccaacgg	cgaggcgctc	12420
gacgtggtgc	tgaactcgct	ctccggcgag	gccattgcgc	gcaacctgcg	cgcgctgcgc	12480
ccgttcggcc	gcttcacgca	gctcgcaag	cgcgactact	acgagaacac	gcacatcggc	12540
ctgcgcgcgt	tcgcgaacaa	catctcgtac	ttcggcgtcg	atgccgacca	gctcatgaag	12600
gagcgggccc	acctcgcgcg	ccagctcttc	aacgagctga	tgcagctctt	cgagcagggc	12660
gtgctctcgc	cgctgcccta	ccgcgccttc	gccgccaccg	aggcggtgga	agccttccgc	12720
tacatgcage	actcgcgcca	gatcggaag	gtggtgctgt	cgcttcgcca	cggcgtgaag	12780
gcgcagcccg	cccgccccgc	cgtgccgcgc	gagctggcgc	tgtcgccgaa	cgccacgtac	12840
ctcgtcacccg	gcggcctgtc	gggcttcgca	ctgcgcaccg	cgagtggtgc	ggtggacaag	12900
ggtgcccgcg	acctggtgct	ggtctccaag	agcggcgcgc	agtcgctcga	gaccaaggcc	12960
gcggtggccg	acctcgaggg	ccaggggcgc	acggtgatcg	cgagggcggt	cgacatcacc	13020
gaccgcgcct	ccgtgcaacg	gctgctcgcc	gaggtcgcg	ccgcgctgcc	gccgctgcgc	13080
ggcgtgatec	acgcggcggc	cgtgatccag	gacggcttca	tcaccaacat	gacggccgcg	13140
cagatccgcg	acgtgctggc	gcccaaggtc	ctcggcgcgc	gccacctgga	cgagctcacg	13200
cgcggcgcga	agctcgatct	cttcgtgctg	tactcgctcg	cgacgacgct	gttcggcaac	13260
ccgggcccag	ccaactacat	cgcgcccaac	tgttctctcg	aggcactcgc	caaggcgcg	13320
cgcgctcgag	gccagccggc	gctgtgcgtg	ggctggggcg	cgatcgggga	cgctggctac	13380
ctcgcgcgcc	acgagaaggt	gaagggaagc	ctgcagacgc	acatgggcgg	caccgcactc	13440
gagtcggagg	ccgcgctcgc	ggtgctggag	cagctcctgc	tcgccgacgc	ctcgggcctc	13500
ggcgtgctcg	acttcagctg	gcgcaacctc	cgccgcttcc	tccccaccgc	cgctcgcgcg	13560
cgcttcgcgc	agctgggcgc	gcgcggcgac	gacgcgaagg	acgacgacga	ccggcgcggc	13620
gagctgcgcc	gcctcgccaa	cgagctggac	gcctccgagc	tctcggcgat	gttcaccgac	13680
ctgctgcgcc	gcgaggtggg	cgagatcctg	cgcacccgcg	cggaaccgct	cgatacgcg	13740
cagcccttgc	aggaaatggg	catggattcg	ctgatgggcg	tggagctgct	gaccgcgtg	13800
gagcgcgct	tcgagctgaa	cctgcccgtg	atgtcgtctc	ccgagcagcc	gtcgatcgag	13860
aagctggctg	accgcatcgt	gcgcgcgctg	aaggaccgca	acgccggcac	cgaggccgaa	13920
tcgcgcagcg	accagatcga	gcgcgtcgcc	gcgcagtagc	cgcccagact	cgactcgcg	13980
caggtcgagg	agctcacgca	ggcggtcgat	gacgcgcaag	ccacgcagca	gggacgtccg	14040
tgaaggcgcg	caacctccgc	ggcctctcgc	aggcagtgaa	ggagcagctg	atccagcgcg	14100
tgctcgagca	gcgcgtgcgc	cggttcgaga	aggacaagcc	ctcggcggtc	gagtcgcgca	14160
tcaacgcctt	ccgccccgac	cccgatgccc	tgggccaact	ctcggagcgc	tactgcccgt	14220
tcgacatgca	cccgtctctc	cagcagatgc	agctggtgaa	ggaaggtgcc	gcgaagctgg	14280
gcattcccga	tcctgacttc	cgcgcgacgc	acggcggtgg	cgcgataacc	acgcgcattg	14340
ccgggcgcga	gtacatcaac	ttctccagct	acaactacct	cggttctctg	ggcgaccgcg	14400
ccgtggatgc	cgcgcccaag	caggccatcg	acgaatacgg	gaccctcggt	ctcggccagc	14460
aggctcgtgt	ccggcgagcg	gccgctgcac	cgcgagctgg	agcgcgccat	cgccgaggtc	14520
tatggcgctg	acgacgccgt	ggccttcac	agcgccacgc	cgaccaacgt	ctccaccatc	14580
ggccacctgt	tcggggccgcg	cgacctcatc	gtgcacgacg	cgcttcgtgca	caacagcatc	14640
ctgatgggca	tccagctctc	gggcgccaag	cgcatggcct	tcccgcaaaa	cgactggcgg	14700
gccctggagc	agctgctcgc	cgcgacgcgc	cgccacttcg	agcgcggtgt	gatcgtgatc	14760
gagggcgctt	acagcatgga	cggggactac	cccagctgct	ccgaattctc	gcgcctgcgc	14820
cgccgccacc	gcgcgttctc	gatggtggat	gaagcgact	ccttcggcgt	gatgggcccg	14880
cgcggtctcg	gcatccgcga	ccacttcggc	atggccggcg	acgaggccga	catctggatg	14940
ggcacgctct	ccaagacgct	cgctcgtgct	ggcggttca	tcgccggcga	gaaggcgctg	15000
gtcgagcacc	tgaagttcgc	cgccccggcg	ttctcttaca	gcgtcggcat	ggccccgccc	15060
gtggcgccgc	cggcgctcgc	cgcgctgaac	cgcatgcgcg	aggagccggc	gcgcgtggcc	15120
gccctgcagg	agggcgagcg	cttcttcttc	gaggccgcgc	gcgcgcggcg	catcgacgtg	15180
ggcctctcgc	ggcgcatggc	cgtggtgccc	gccatcaccg	gcagctcgat	ccgcgcggcg	15240
cgcttgccgc	aggcgatggt	ccagcgcggc	atcaacgtcc	agccgatcgt	ctaccgggcc	15300

28/39

gtgcaggaga	actcggcgcg	cctgcgcttc	ttcgtgagcg	ccacgcacac	cgaggagcag	15360
ctgcgcttca	cgggccgcga	gctcgcgcga	gcctggcgca	agctctgagt	ggcgggcccc	15420
cggctgcgcg	tcgtcgtctg	cacccacggc	ggcctgaacg	gcgcgctggt	gctctcgcg	15480
ctgctggccg	ccccacgct	cgaagtctcc	gcgctggtga	tctccagccg	cgcgcgcggt	15540
gccccatgagt	cgtatggccg	tgcgcgcctc	ggatacgtgc	gcgcgagcgg	cgtcgcctat	15600
gcgctgtacc	tctggtgcgc	cacggcgctc	gccgacctgc	tgctgcgcgg	cacgtccgag	15660
ggccccggtcg	cgcgcatcgc	cctcgcgcgc	ggcatcccgc	tcctcgccac	gccgcgcgtg	15720
aacgatgcca	ccgcgcgcgc	cttcacgcgc	ggggcgggcg	cggacctgat	cgtctccgcg	15780
ttcttcaacc	agcacatcga	cgccgacgtg	gctgcgctcg	cgcgctggc	cgcggtcaac	15840
atccaccctg	cgccgctgcc	gcacttcgcg	ggcgtggatc	cgggtgcctt	cgcgctcctg	15900
cgcggtgccg	agcgccacgg	cgtgagcggtg	catcgcatcg	aaccgcgctt	cgataccggc	15960
gcgctgctcg	cccaggaaac	cgacgtcgag	gccccggca	gcgtctttgc	cgccaccgcg	16020
gcgctctatg	accgcggggc	ggcgctcctc	gccggcggtg	ccgcggcgct	ggccgcccag	16080
ccgcgcggaa	ccccccagcc	ggcggggggc	tcctacgact	cctggcccac	ccgcgcccag	16140
gtcgccgcct	tcggcgcgcg	cgcgggggcg	ctgctccgcg	cccgcgacct	gtggcgccctc	16200
gcgcgcgggg	ggccggccgc	tttcgtaata	gaatcagcgc	ggtagcgggc	actcacgcgc	16260
cgcttttccc	cgaccccccg	aggccccatg	aacactggct	gaagcaacac	gcaatctcgc	16320
cgctcaccgt	accgtgccc	cgatcctcta	cttcggcctg	atcgccctcg	atcgccctcg	16380
acgtctacat	ctccgaatca	cgcttcgtgg	tgaggagccc	ccagcgccag	gtgcagaccg	16440
gcctggtggg	cgccctgctc	tcgggcaccg	gcttctcgcg	ctcccaggac	gacacctact	16500
cggtgcacga	cttcacaccc	tcgcgcgacg	cgctggcgga	gctggacaag	aagctcgccg	16560
tccgcaagct	ctacacggcc	gccaacatcg	acttcacaa	ccgcttccc	gggctcgact	16620
gggacgacag	cttcgaggcg	ttccaccgct	actaccagaa	gaaggtcacg	atcgacttcg	16680
acaccgcgtc	ctcgatcacg	gtgctgcgcg	tgcgcgccct	cgagaaggcc	gactcgcggc	16740
gcatcaacga	cctgctgctg	cagatggggc	agcgctggt	gaacgagctg	aacgagcgca	16800
ggcgccagga	cctgatccgc	ttcgcgacgg	ccgaggtctc	gctcgccgag	gacaagggtga	16860
aggatgcgcg	gctggcgctc	tcgcgcttcc	gcagcaacca	gtcgggtgttc	gagcccagacc	16920
gccaggccctc	gatccagctg	cagggcggtg	ccaagctcca	ggaggagctc	atcgccaccg	16980
aggggcagct	ggcacagctg	cgcaagctct	cgcccgacaa	cccgcagatc	ggcgcgctcg	17040
agaacaagtc	ggcgcgctg	cgcggtggcg	tgcgcgcgga	gtccgcgaag	gtgaccggcg	17100
gcagcggtc	gttcagcgcg	cgcgcccccg	cgttcgagcg	cctcacccctc	gagaagggtc	17160
tcgccgaccg	ccagctgggc	gttgccctca	ccgcgctgga	gaccgcgcgc	agcgaggcg	17220
agagggaaga	gctctacctc	gagcgcatcg	tgcaagccaa	cctgcccgcg	gaatcggtcg	17280
agccgcgcgc	catccgctcg	atcttcaccg	tggtcgtgct	gggcctggtg	gcctggggcg	17340
tggtgagcct	gctggtggcg	agcgtgcgcg	agcacgtcga	ctaggccgtg	gagcagaacc	17400
cttccctcct	gcgcgcctc	ggggtgcagc	ggcgcggtg	ctacgcgctg	ctgatgcgcg	17460
aggatcatcac	gcgcttcggg	cgcgacgacc	tcggcggtg	gtggctggtg	gtcgagccga	17520
tgatcttcac	gctcggcggtg	acggccctgt	ggaccgcggc	cggcatgaac	cacggctcgt	17580
cgctgccgat	cgtggccctc	gcggtcaccg	gatactccgc	ggtgctggtg	tggcgcaact	17640
gcgcctcgcg	cgccctgatg	gccatcgagg	ccaacaagg	cctgctcttc	caccgcaacg	17700
tgcgctgat	cgacgtgttc	gtgacgcgca	tcatacctcga	gatcgcgggc	gccacggcgt	17760
cgttcacctg	gctgggcac	ttcttcacat	ggatcggtg	gatgccgatg	ccgatcgaca	17820
tgctcaaggt	cgcgcttcggc	tggttcacat	tgctcgtggt	cggcgcccagc	ctggcgctcg	17880
ccatcggcgc	cggcaccgcg	tactcgctgg	tggtcgagcg	gctgtggccg	ccgacggcct	17940
acctgctgtt	cccgctctcg	ggcgccgcct	tcattggtcga	gtggctgccc	gagaaattcc	18000
agcagttcgt	gctcttgctg	ccgatggtgc	acggcaccga	gctgctgcgc	gacggctact	18060
tcggcaacgc	ggtgcgcacg	cactacgacg	tgggctacat	ggcaacggtc	tcgctgggccc	18120
tgctcgtgct	cgcatgcac	ctggtgcgca	tgccggccaa	gcgcgtggag	gccgcgtgat	18180
ccgggtcgag	aacgtgacca	gatgtactca	cgcgccctcg	cccagaacgt	gctgacgcgt	18240
cagcttcgag	ctcgcccgcg	ggcgcaacat	cggcacctc	ggcaagaacg	gcgcccggcaa	18300
gtccacgctg	atccgcctga	tcagcggcgc	cgagagcccc	accaccgggc	gcatcgctgcg	18360
cgagatgagc	gtgtcctggc	cgctcgccct	cgcgggcgcg	ttccagaccc	acctcaccgg	18420
gctggacaac	ctgaagttcg	tctgccgcac	ctacggcgctc	gactaccgcg	acaaggtcgc	18480
cttcgctcag	gacttcaccg	agctcggcgt	gtatttcgcg	gagccgggtg	tccactactc	18540
gcacggcatg	atcacgcgc	tcgccttcgc	gctgtcgatg	gcggtggagt	tcgactgctt	18600
cctgatcgac	gagggcatgg	tggtggcgga	cgcgcgcttc	cacgagcgct	gccacgtcga	18660
gctcttccac	aagcgcaag	accgcgcctt	catcctcgtc	acccacgacg	ccaaggtgat	18720

29/39

caagctctat	tgcgagagcg	cctgcgtgct	gcacgagggg	cggctgctgc	ccttccccac	18780
cgtggacgcc	gacctacgag	tctacatgaa	cgaggtcatg	caggacctcg	ccccggaggt	18840
tgctgaacg	ccccgcgcg	gccgctgcgc	gtgctcctcg	agctgcggcc	ggcgttcgac	18900
ggccacgccg	gcatcccgca	ggaggcacgc	ctgctgttcc	gcggcctgcg	catgatcgag	18960
ggcatcgacg	tggagggcct	gctgcagcac	agcggccacg	tgctcgccaa	gggctgccg	19020
ccgcgcggcg	gcggcgacct	cgcaccggac	cgccagctga	accgcctgtc	gcgcgtcgtg	19080
gtctccacca	agcaggagct	caccaacgcg	cacgtggcga	cggcgctgat	ggccattcgc	19140
cagctgctgg	gcgcgcgcga	gaccctctac	cgcttcgacg	cggcgcaactt	ccgcgacttc	19200
gtctggcagg	cgctcttcgc	gcgcacgctg	cacgccgcgg	acttcgactc	ggtcaccgcg	19260
gccggctttc	gcgtcgcgcg	cgtgccgtgg	acgggcatgc	accgctgcgc	gctggtcacg	19320
cgcaagctcg	gctacacgct	gcatccgcgc	atcgacactt	ccgatttcga	ctggtcgatc	19380
gccgagacgc	cctacccggc	gcgggtcgcg	gacggcacgc	gcctcgtggg	gcgctaccac	19440
gacgcgatcc	cgctcctcat	gccgcacacc	atctccgaca	tgtcctacca	ccaggcctcg	19500
cactaccggg	cgctgcgccg	caacgtggcc	tcgggcgcgc	acttcgcgtg	cgtctccgaa	19560
gccacgcgca	aggacctgct	ctcggctctt	cccagggtcg	aggcgcgctc	cagcaccatc	19620
cacaacatgg	tctcgcacca	ctacttccgc	gagaccgccg	acgccggggc	catcgaggag	19680
atcctgcgaa	cgcgcgcgag	cgagcgcatc	aaggcgcgcg	ccaacggagc	cgccaccgcc	19740
gtcgcacgat	cgccctcctt	cctcgcgcgc	gccgcgcaac	ccagcgaccc	cggctaccgt	19800
ctcgcggtgg	cgacgatcga	gccgcgcaag	aaccacgcgc	gactgatcgc	cgcgtgggag	19860
cacctgcgca	cctcgcgctt	cccgggctcg	cgcttggtgg	tggctcgcat	gcccggctgg	19920
ggctacgagg	cgctggtagc	caagttcaag	ccgtggctcg	cgcgcgccca	gctcttcgtg	19980
ctcgaggacg	tgcccgcgcc	cgagctgcgc	ctgctctact	cgacgcgcgc	tgccacggtg	20040
tgcccagagct	tcggcgaggg	cttcgatttc	tccggcgtgg	aggcgatgcg	ctcgggcagc	20100
ccggtgatcg	cctcggagat	cgccgcgcac	cgcgaggtgt	accgcgacgc	cgccgagtac	20160
tgcagcccct	attccgtggc	cgacctcgcc	gaggcgatcg	ggcgcgatcg	cgaccggcg	20220
gcgacggggc	tgcgccaggc	gctggtcacg	cgcggcaccg	aggtctcgca	gcggtacacg	20280
cccgaagcca	tcctgccgca	gtggcgcgag	tacctgctcg	gcacggtgcg	cgcgagggcc	20340
ccgtgagcga	tccgaaggcc	accgaccagg	cgtgggagga	gtggggccag	cgcgaccctt	20400
acttcggcgt	gatcaccaac	ccgcgcttcc	gccgcgggca	gatggacgag	gaagcgaagc	20460
gcgaattcct	cgcgctccgg	cgcgctccat	ccgactacgt	gatgcgcatg	gtgcacgcgc	20520
acatcgcgcc	cgacttccgt	ccgcgcacca	tcctcgattt	cggctgcggc	gtgggcccgc	20580
tcgtcatccc	cttcgccgcc	caggccgagc	aggtgaccgg	cgccgacgtc	tccccctcgt	20640
tgctggccga	ggctgcgcgc	aactgcgcgc	agcaggggct	ggccaatgcg	cgctcgtgg	20700
tctcggacga	ctcgctcacc	gggctgccgg	gccccttcga	cctcgtgcac	tcgttcacgc	20760
tgttccagca	catcgatccg	gcgcgcgggc	gcgacatctt	ccgcgccttg	ctgggcaccc	20820
tggcgagcgg	cgcgctgggc	gcgttgcact	tcgtctatgc	gaagcgtatc	tacgcggcga	20880
cgtacggcgt	ggcgccaccg	cccagaccgc	cgccgcccc	gccgcgcgcc	cagcctccac	20940
ccagccgggc	ggagatcaag	gccgcgagca	aggcgcgggc	cgcgctggcc	acgcaggcac	21000
ggcgccggga	accggcaccc	aatcccgaac	ccgacatgca	gatgaattcc	taccggcgcg	21060
cggagatgct	gttcctcgtg	caggaagccg	gcgtcacgcg	cttcacgtc	gagttcaccg	21120
accacggcgg	cgagctgggt	ctcttccctt	tcttcgcgaa	gccctaggcg	tccttcgagg	21180
cgaaggagag	gcccagcgcg	gtgatcgtcg	cggcgtgctg	cggggcgacc	gccgcgcgcg	21240
ccgcgacttc	ggcagccacg	cgctcgaggt	cgccgcgcga	cccgtacagg	cgcacgcgcc	21300
agtagtcac	ccagtaggag	tccggacagg	ccgacgcgat	tcccagcgcg	cgctcgacct	21360
cgggcccgcg	gcgcctcctg	tgcacttggt	gcgaggccca	ttgcaggttc	acggagggct	21420
cgaggggcag	gacggcgagc	agcgcttcgc	tcaccgccag	cgcgcgcgcc	tcacgccgct	21480
cgtcggcgag	gcgcatcagc	gtgggcaggg	actcggcgcg	aagcgcggtg	gcggcccgca	21540
gcgaatcgcg	cagcatcgca	gcgtcgcgcg	gcgcgagcgc	caggcgggcg	cggtaggtgt	21600
gggtccagaa	cgaaggcgcc	gcggcatcga	gcgcgcgcgc	gacctcggcg	cttgcgtcgc	21660
ccccggagag	ccgggacgcg	gcccacaaca	ggttcgcctc	cgccgcgtgc	gccggcacgc	21720
tccgcaacgg	cgccagcacc	gcggtggcga	cctcgcgcga	accctcgcgc	gacgagaagg	21780
ccgagcgcgc	gaactgcata	agcgattcgc	cgagcgacga	atgttcggcg	cagagccgcg	21840
cgggatccac	cgcgccgagc	gcgcgggcgg	cacgcagcgc	cgctcctcgc	tcaccgcgaa	21900
gcgcgtgcag	ccgcaggcga	tggaaacgcg	cccactccgc	cgacatcgac	gccgccccgc	21960
gctccagcgc	accctcgacg	gcgccatcgc	tgcgccgcgc	cgcgatgtgc	aggatggccc	22020
aggtcgcctg	cgcgccgggg	aggttcgtgc	cttcgcgcac	gagcggtgcg	aggaccgcga	22080
gcgccgcggc	ggcgctcgcg	gcgccactga	agtcgttgcc	caggccgatc	agcgccccgc	22140

30/39

cgagctcggg	ggccgcgcgg	aaggcggccg	ccgcgtcgcg	ccccaccgcc	gcgaggtcgc	22200
cgcgacggcg	gtagaggcgc	gcgcggtggt	aatagcccca	gaacgggtcg	gggctcgcga	22260
tcagtgaatc	gaacgagcgc	togacctcgc	cgcccacgtt	gccgaggtcc	acctggtgcg	22320
cgccccacag	ggtcgcccga	tgcgcggcca	tcggctcgtc	ggggacgaag	gcgcggtgga	22380
tcagctcgac	gtcgccgaag	cgccgctcgc	gcaacgcgag	ctccagggcg	ttcagccgct	22440
cgaggtcgcg	cttgtagctg	cgataggcat	cgcggtcgcg	gtcgtattcc	tgcgcggcac	22500
tgtcgtccat	cagcgaggcc	gtgagccgct	ggtcgccgtg	aacggtgtcg	tatgcgtagg	22560
ccgcgcgac	gtcgtagggc	agcgtggccg	cgcgacgcga	cccgaacgcg	gcgcggtcgc	22620
ggaatcccag	cttgcccagc	acgcggggtg	ccaggtcgtg	ggcttccatc	tcgcccgggc	22680
tcagctcgcg	cagcgccggc	acgttcgagt	agagccacaa	cggcgtgcgg	tagagctgcg	22740
ccaggtccgc	gtccggggcg	ccctgcgcct	cgaagatcgc	cttccactcc	gcgagtcctt	22800
cgtggccggc	ggagaagatc	ggcacgtggt	cgccgaagat	cgccagcacg	tagggccgcc	22860
cgcgacacct	gagacgctcg	accagcgagc	gcagcgcgcg	gtcggtttcg	gccatcaccg	22920
cgcagtaact	gtcgtatcgc	ccctggcgca	cgtgcccggc	gaggaagtgc	ggcgccacgt	22980
agcgcggtc	ggcgcccttg	tcgcccaggt	agggtcctcg	gcccatacgc	gagacgccga	23040
acacgaagcg	ggcgccgcaa	tcgtcgagct	cgcgatcac	ggcatccatc	acggccgaat	23100
ccgcgatgcg	ctcgaactgg	cgcgccgcat	gggcgaacgc	ctcgtcggcg	tggaaacccg	23160
cgaagccgag	cttcttgtag	acgcggtcgc	ggttgtagaa	ccagcggtgg	aacggatgga	23220
tggccacgct	gcggtagccg	tggcgccgca	gcgtgcgcgg	caccgcatac	acccggtcgc	23280
gcaactgcc	catgtagggc	agcgcgccgc	agccttcctc	gcccataatgc	atgcccgtga	23340
ggatctcgaa	ctcgtgtgct	caggtgttgc	cgccgatcac	cttcacgcgc	agcgccggcg	23400
tgggtcggcg	cgccgcgctc	gcgcgagga	acggcagcgg	gtcgcgcgcg	agcgccacgc	23460
cgaacgggcg	caggtcggcg	agcgactcca	tcatacaggaa	gatgatgtcg	ggcggtcccg	23520
cggtggccgg	gccgctcgcc	gagcgctcct	ggcgctccag	cagccgctcc	acgccctcgc	23580
gcagccgcgg	cgccgcgacc	ggcgccggcg	tgctcccggc	gaccagcacg	gcgtgcagga	23640
cgaagccacg	gcgcgcgacg	ttcacgttgt	ggtcgggtgg	cgaatacgtc	atcgctcgca	23700
actccaccgg	ccagcgccgc	aagcgtccgg	cgacgaaggc	gacgccgagc	gcggcgagca	23760
cgagcagcca	ctgcggcgca	tcgtcgtacg	cgctcgtcag	gtcgtatggc	tcggtgatct	23820
cgaacgcggc	ccacgccagc	gccgcgagca	ctgcggcagc	ggcaagcgcg	cgcaccgcca	23880
cgcgcgggcc	atgcaggtac	gccggcgcgca	acaccagcgc	ctcgcgagcg	ctcaggaaat	23940
cgccccggcc	gaccacgcgg	ccgaagaacg	cgagcaggcc	tgctgttacc	gcgtagagcg	24000
catggaagat	ggcgagcagg	ccggcgcttc	agggcaggcg	ccgcgcgcgc	gcccaggggt	24060
ccacgtcata	gcccaccacg	aggcgccgct	cgagggcggc	ctgcagcacc	gcgtaggcgca	24120
gcacgcagac	cagcgtgccc	cccagcagca	cgccgaggaa	acccgccaaa	ccagccgtgt	24180
cgtcgaggat	catgggacca	ccagggacgt	cttcggcgca	gtctataacg	ccccgccccca	24240
tgaaaaacgg	cgaggccccg	gggcctcgcc	gtggagggaa	ccgcgcaacc	tcagggttgc	24300
gggggtgcga	ccgaccagcc	gccggggccg	atcaggtttg	cgctgccgat	ggtggcgagc	24360
ccgttcacga	ggcgacccgt	gacgtgcgcg	tcggtgtggc	ggaagaccag	gtccttgccg	24420
ctgtcgccgt	tcaggtccag	cagctgcgtc	acgttccaac	ccgtgcccgc	cggcagcgca	24480
tcgcccgcgg	tgatgatgga	aacaccatcc	atcaggcgga	tgtgcacgcg	gccgtcgggtg	24540
tggcggaaga	ccaggtcgtc	cttcgcgctc	ccgttcagg	cgcccaccag	gttcaccgac	24600
caccgcgtgc	cgccggcgag	gagctggggc	ctggcgccga	aggtgggtgc	gttcacacagg	24660
aacaggtgcg	cgcgccgctc	ggtgtggcgg	aagatcatgt	ccgcgcgctc	gtcgcggttg	24720
aggtcgccca	ggtggcttac	cgtccagccg	ctggcgcccg	agaggaagcc	cgctccggcc	24780
gtcaccgtgg	tgcggtccat	gatgtagatg	tagccgcggc	cgctcgctgtg	catgaagacc	24840
aggtccgcct	tgcggtccgc	gttcagtgcg	ccgcctggg	tgagcgtcca	gcccgtcccc	24900
gccccgaaga	gctgggcgct	gccgatgatc	gtcgtgccc	ccatcaacca	gatgtgcgcg	24960
cgtccgtcgg	tgtggcgag	gatgatgtcc	gccttgccgt	cgccgttcag	gtccgcccgtg	25020
tggctcacga	tcagcccgag	cccggcgggc	agcagctcct	tgccggccgt	caccgtagtc	25080
ccgttcacga	tgtacgcgta	ggcgccggca	tcggtgtggc	ggaacaggat	gtcggcgccg	25140
aggtcgccat	tgaggtcggc	cgtctggctc	accgtccagc	ccgcgcggcg	gccgatcagg	25200
ttggccgagc	cggtgatggc	ggtggcgttc	atcgtccaca	ccgcgatgcg	cccgtcgctg	25260
ttctgcagca	cagagtcgct	cctgccgtcg	ccgtccaggt	ccgagacgac	cggttggtg	25320
ggcgccgctga	cgccaccgga	aaccggcagt	gccagcccgt	cgatccaggc	gcgggtcggcg	25380
cagttggcgc	catccggggc	cgcggggttc	caggtcggcg	agttgcaagg	cgtggagagg	25440
aagttcgaga	aacgccacac	cagcgtgtgc	gcgcccgcgg	tcaccgggaa	cgacaccagc	25500
						25560

31/39

ttccagcccg	aggccgtggt	gcccgcgtcg	ctgaacacca	ccgtgccgtc	gatcaggaac	25620
tcgaacttgc	cggcggttggg	gaagctcgac	acgcgatagg	cgaacgcgac	gttgcccgcg	25680
agcagcgtgc	ccgcgaacga	gaggctcgag	ttcaccgtcg	tcgagttggt	cgggtcgctc	25740
gtcaccacct	gcgcggaacg	aaggctggtc	gcgccctcga	aggcctggtc	ggagccgacc	25800
gtccatgcgt	tggcgccgcc	cgaggtcggt	aagccggcag	gcagcgtgcc	accctggggc	25860
cacggcgaga	ggatgatgcc	ggtcgcggtc	gccggggccg	cgatgaagac	accctgggaa	25920
cccgtcgcat	tggagagcgc	gacgttgaac	gtctcggttc	cttcaccgat	accgtcgctc	25980
gccaccggga	ccgtgatggt	cttggggcgtg	gtctcgccgt	tggcccagct	gagcgagccc	26040
gaggtcgcg	tgtagtcggc	gcccgaagt	gcggtgccgt	tggaggtggc	gtagctgacc	26100
gagatggcac	ccgaccgagcc	gccgatgcgg	ctcacgggtga	gcgtcacgtt	gccggcggtt	26160
tccgcggccg	cgaaggtggt	gccggtgaac	tgcaccgaac	cgggaacggc	gggcgggggtg	26220
accgtggact	tgagcgccga	gagtgccgcg	cggttggtgt	tcagcgagag	cgctcgcttg	26280
gcctgcgcgg	tgcgcacagg	ttgcggcgca	ccgacacccg	tcgtgcacga	gaggcccggg	26340
ttggagaagc	ggtagacctt	ggtcgctcag	tcatgcagca	gcgacatgat	ggtggtgaag	26400
ttcgccgcgt	cgttcgctcg	gcactcgggg	cgctggctca	cgccgcaggc	gccgcccgccg	26460
gagccgttgg	gcagggccga	gttgcaggtg	aggttgccgc	tcgcgcaata	gaagtggccg	26520
tacgagtagc	cgaacgcacc	ttgcggcggg	ctcgcggtgc	cgcccgcttc	ccaggcctgc	26580
gagggcggtg	cgtggcggtt	gcccattggga	tggccgagct	cgtggatgaa	gacccactcg	26640
caaccgaaca	ggcagccgga	agcaacggcg	taggcaaagt	tggggtccgg	cgtgctggtg	26700
ccgatccagg	ctgcgcgctt	gccgccgaag	tcggcgccgt	tgcgcatgaa	ggccaccatg	26760
tccgcgcgct	attgcgtgcg	gatggcctcg	atgttgccga	acgtcgccgg	gtcgaagctc	26820
gcgtcgcccg	gggtgatggc	atgcacgcgc	gtggtgtcgt	cgatggcgtc	gtcgtagtct	26880
acctgcgtcg	cattcaccag	gcgcagcgtg	atggcgacct	cgctgtcggc	atacgcagtg	26940
ttggcgcgcg	tgacgaggaa	gttcaggcgc	gtcatcaggc	tgcgcgccgag	gcgctggccg	27000
aagccgcgcg	agtacacgat	catcaggctcg	atgacgttct	gcggcggtgg	cgtcacctgg	27060
gcgagccgac	ggtattcacg	ccggggatgg	cgatcagggt	ctcggggctg	ccggtcgcgga	27120
cgacgcgcga	cttcgagccg	gtcgtggcca	cggcgtggtc	ggcgccgagg	ctgatgatcg	27180
gcatgctggc	ctgctcggcc	gtcatgtcga	ccaggtagtc	gacggtgccg	cccggaatca	27240
ggcgggtactc	accctgcggg	gtcgagaaca	cgccgaagga	gccttcaggc	ccggtcgtga	27300
cgatcgcgcg	gtggttgatg	ccgtcggtct	tgtctctcgc	gatccacgac	gtgatgccgt	27360
cgccgtgggc	ctggagcaac	tcgaagacat	acgagtggcg	caccccgttg	gggagcgaca	27420
gctcgacctc	ggagcgcgga	ctcagggcac	ggagggccgc	ggcgttgaag	cgcaccgctt	27480
gccgggcaac	ggcgctgccc	ggcgccggca	tcgcggccgc	gggtgcggag	accaggatct	27540
ggggaaccgc	ggcgaaggcc	acggaagaga	agacgatcgc	cagcgcgccg	aggaaggcgg	27600
ctgtgatccg	ttgcgtgaat	gcgttcattg	ggagctccgg	aagttgacct	atgcccaatc	27660
cgctatgtcg	cggagatgtg	gacaaaaggt	atcgagcggg	cgtgacgacc	cgcccccgga	27720
gggatgctcc	aaaaggacta	cgaggggtgct	acggctgggt	tgggtggcgaa	ggccgaagga	27780
gaactccttg	tgtctccgc	gcgacttaag	gttgcgaggg	aaccaccgac	caccaccggc	27840
ggccgatcaa	gtccgcgctg	ccgatcgtgg	tgacgcgctt	catcaggcga	atcgtcacgc	27900
gcccacgtt	gtggtggaag	accacgtcca	tgcgcccgtc	gccgttgaag	tcgaggagct	27960
gcgtcacctg	ccagcccgcg	aacggcggca	ggatgtcggc	ggcgccgagg	atcgcggtgc	28020
cgttcatcag	gcgcacgtgc	gcgcggccgt	cggtgtgcgc	gaacacgagg	tccgcccggc	28080
cgtcgccggt	gagatctccc	accaggttca	ccgaccatcc	ggtgcccgcg	ggcaacagct	28140
cggcgctggc	gccgaacgtc	gtgccatcca	tcaggaaacag	gtgtgcgcgg	ccgtcgcttg	28200
ggcggaagac	gaggtcgggc	ttctggtcgc	cgctgaagtc	cgcgacgtgg	ctcaccgccc	28260
agccgctgcc	ggcgcgagg	aagcccacgc	cggcggtgat	cgccgtgccg	ttcatgatgt	28320
atgcgtagcc	gcgcccacgc	acgttggaag	agacgatgtc	ggccttgccg	tcgccgttca	28380
tgtcgccggt	gccgaccacg	ttccatcccc	agcccgcggg	cagcaggctc	gcactgccgc	28440
tgatcgcggt	gccattcatc	agccagatgt	gggcgcgtcc	gtccgtgtgc	tgagcagcag	28500
gtcggccttg	ccgtcgccgt	tcaggtccgc	cggtgtggctg	atggtccagc	ccgtgccggc	28560
ggggaacagc	tccttgccgc	ccacgaccgt	gagcccgttc	atctggtacg	tgtagatgcg	28620
cccgtcggtg	tgttggaagt	agatgtcggc	ctggctgtcg	ccgttggaat	ccgcgcccat	28680
tgatcacgct	ccagccggcg	ccggcgggaa	tcaggttcgc	cgaaccagtg	atcgtcgttc	28740
cgttcatggt	ccagcgggcg	atgcgcccat	cggtgtgctg	gaacagggaag	tcgctctgcc	28800
cgtcgccggt	caggtcggtc	gccgcatggg	tcgggaacgc	caccttcgcy	aggaaggcat	28860
cggagaaacc	cgtgagcaac	gtcttgtacg	cgccgggcgt	ggtcggatag	ttcgccgacg	28920
acgtgcggcc	ggccacatag	gccgcgcccc	gcgcacgcac	gccgacgccg	tacccacact	28980

32/39

cgacgttgtt	gccgcccgatc	agcgtcagaga	acacgatgtc	gccgttgcgc	acgcgagtga	29040
ggaacgcgtc	ggcatcgccg	ccggggccct	gcaccgggtt	cagggtcggg	aagccggcga	29100
cctgcgtgta	gccacgcacg	agcgcacgc	cattcggcgc	cagggcgacg	tccaggggat	29160
tgtcgtaggc	cgagccgccg	aagatgccc	acgattcgag	cgcccggtg	gccgagaagc	29220
gcgtcacgaa	cgcatcgacg	gtgccggcga	aattgcgcac	cgttccggtc	tgcgggaagc	29280
tgggcaggcc	ggtatcgccg	acgacggtcg	cctggcccgt	gaccgggttc	accgcgatcg	29340
cggtcacgtt	gtcgccgctg	ccgccgagga	acgtggaata	cgcgaccggg	ccgccgctcg	29400
caccgatctt	ggtgacgaag	ccgccaccgg	gttgggtctg	gtacgcattc	acgttgcgga	29460
acgtcgagga	gcgcgtggag	cccgcgacgt	acgtgttgcc	cgcatgtgcc	gttgcgatac	29520
cgcgccgat	gtcatcacac	ctgccgccga	gatacgtgga	atagacgaag	ccgccgttga	29580
acgagagctg	gtgacgaacg	cgtcgcgatt	gaggccggtg	ccgacgaaca	cggtgcagc	29640
gcgtgccgc	cggtcgggaa	gccgcgcgcg	gccgaaccgg	tgacatgcat	gaagccgttg	29700
ttgtccacgg	cgatcgccat	gccctcgctg	tcgccggagc	cgcccagcag	ccgcgagaag	29760
atgacgttgc	cagtgggtatt	cagcgccgtg	acgaatgcat	cgcgcccggc	ggcatggttg	29820
gtgcaatcgg	tgacggcgaa	gcaatgcggg	aacgcgttgg	tgccggtgga	accggtgatc	29880
gcgacgaact	cgttggccgc	ggaggcgccg	agcgccagcc	cgcgagata	cgcttcggcc	29940
ccgagcgtgg	tcgagtacac	cacctgctgc	gtggtggggt	tgaacttcac	caccgtggct	30000
tcgtacgtgc	ctgcgggtata	ggtcgcgagc	ccgacgtaca	cgttgccttg	cgctccacc	30060
ttgaccggcg	tgggtccactg	ctcgtcgccc	gcgccgccca	ggtacgtgga	gaacgccatc	30120
accgggtcga	tcaccagcgt	gcgggtgctg	tcgtagtcgg	cgatcacgaa	gccggcctcg	30180
gcgccctgct	cgcccacgaa	gagctcgaag	cgcgcgccca	ccggcacgcg	cgaatcgccg	30240
acctgctgga	aggctaccgg	cgctgctgc	gtaaatccct	cctggcccac	gcgcatgcgc	30300
aggttgcctt	cgccatcgat	ccacgcgtgg	tcggcgagcg	agaggtcgag	gcggtatctg	30360
cgcggtacgg	cgcgcggggc	gaccacgaag	tcgtattcga	gcgtgccctc	cttgcctgac	30420
acggtgaggt	ccacgcgcg	atacaggctc	ttcagcgaca	cgcgccgaa	gtcgggcacg	30480
ttttcgcgcc	ggccggcgcc	cgtggtgccg	ctgtaatagt	ggctgagggg	ggcccggggc	30540
tcctcgccct	cgatgaccgg	atcggaagcc	cccgcgaagc	gcacgcgcaa	cagcgaggcg	30600
ctggcagctt	tcctggagcg	tggctcgaag	gcaatgccat	cgcggtcac	ggagacgcgt	30660
ccggcctgcc	cgcgcgagac	gtagagcgca	tcggggccga	actggccctc	gttgcgctcg	30720
aacgtgatgg	gcgcgttcct	gagggcatca	gggacgacgg	catggacggc	ggcgggagtg	30780
gcgagggcgg	ccgcgacgag	cacggccagg	ggcgggagcg	ccgggcgggc	gccgtggag	30840
ctgcgggtca	tggggcagtg	acctttctgt	tgttatggct	gcttgttgc	tttgcgggat	30900
ccacagctcg	gatccagctg	gcgatactac	cagcgagcga	gtggtttttc	gtgacctagt	30960
cccgcagtcg	ggccccggcg	acaacggcgg	catcgcggtc	ctttgccagg	gcacgcacgg	31020
cctcgatcca	gtcgccgtcg	cgggccagca	cgcggggggc	atcgcggtag	ggctcgagat	31080
ccgaggccac	caccgggatg	ccgaggcagc	cgtactcgag	cagcttcagg	ttgctcttcg	31140
cgcgattgaa	cgggttgctg	cgcagcgggt	ccaccgcgac	gtcgagtgcc	agcgacgcga	31200
gcttctccgg	gtactgcgcg	atgggcacca	tgtcgtgcac	ttccgcggcg	aacggcgcca	31260
gctcgggctg	gcacaggccg	aggaacacct	agtcgatctc	gcgatgcgtg	gcgcgcacca	31320
cgggctcgag	caacttcagg	tcttccccat	gctgtttcgc	ccccgccag	cccacgcgcg	31380
gacgcgcgcc	acccgcggga	cggttcgcga	ggccggccca	acgtcccgca	tcgatcgctg	31440
tcgggatcac	gcgcacgtcc	ttcgcgccgc	ggccgaaggc	ctccgcgagc	ggtgcgggtg	31500
acaccacgag	gcggtcgcac	aacgccaccg	cgcgcgcgat	gcgctgggcg	atgtccgggt	31560
agatcgctcg	cgcatacgga	ttgccggggc	gcagttgcgt	gagcaggtca	tccaggccca	31620
ggaccttcag	cgcgctccg	tggcgcgcg	gcacctcgag	cgaggtgagc	tggtagttgt	31680
ggaagaagtt	gtgcgcgagc	acggcgtcgg	cgtcgaggcg	ctgccactcg	acgcggttgg	31740
gcgcgcagcc	cgtggcgtgc	tcgcccatca	tcacgacctc	gacgcggccc	gcgcgtcca	31800
gcgcgcgcca	cggctggcgc	acgcgcacct	cgccggagcc	ccagcggtcg	aacggaaacg	31860
cgcagagccg	aaccgcgtgc	gtgtcgttgc	cgcgcgccgc	gaagcgttcc	acgggcccgt	31920
agccctcgcc	gcccattgcgc	aacgcggggt	ggtagtgcgg	atcgctcgtc	agcaccggct	31980
gccagcgctc	gcgcattccac	gcctcttcgg	aaggaatgac	gagcttcgcc	accggaacga	32040
cggcatcggc	gcgcgcacgc	gcgagcaacg	ggaagtccgc	cgcgatgccg	gggcgcgcaa	32100
ggatgtcgaa	gcccgcgtgc	cgcagcccga	ggcacaggtg	cgccatcgcg	aacggcccgc	32160
cgcgctcgat	ctcgtgcagc	gcctgcgcgg	agagcgccgc	gtcgcggttg	accagcgaga	32220
gacggcccgc	cgcgacgctc	acctcgcgcg	gcgatgcgta	gagtgccgcg	aggcgctcct	32280
cgccctgcgc	gcgcggcgga	gcgcccagga	cggcccacgg	cccgcgcgcc	gcgatttccc	32340
agccgggaac	gcgcacaccg	ccggggccca	gcaggtcggg	cgcgagggca	cccgtaggag	32400



33/39

ggcctgcat	gccctgctcg	aggcgctcga	gccagccggc	gcggaattcg	ctgcaacgcg	32460
catcgacgat	cgcgatccac	agcgctcgccg	cggttgccgat	ggcgcgcgcg	aacgcggagc	32520
cggtttcgcc	gccacccgga	accaggtgcg	tgcgcacgcc	gcgcccgcgc	agctcaccgg	32580
gccagcgttt	cggcgcctt	cgcgctcgaga	tgcacgtaca	cgtggccgat	acgttcgcgg	32640
cagcccttca	cgagcgcctc	gatgcacgcc	accagccgct	ccggcgcggtg	cgtcgcgtca	32700
atgaaggcgc	aaatcgtgcc	ggccccgatc	gcgcgcgcgc	acgatgcggg	cgcatcgga	32760
gcagctggca	ggaacgcggc	ctgcgtgccc	gtgcggcgga	agtgtccgc	gaccacgcgc	32820
tgggtggcgt	cgcgcgcgcc	gtcgctttcc	tgcgcggcga	tggtcaggtc	gaggcgatgc	32880
acgagcagct	cgcgcgagtc	caccacggcg	ctcgcacggc	cctcctccgc	gaggcgagc	32940
gcgagatcgg	tgacgcccgc	ccagcccgcg	gcgcgcagcc	cgccacgctc	cagcaacgca	33000
gcgcggcgca	gcaccgcgag	cccggcgagc	ggctggtgtc	gcgcgaactc	cgcgccgaac	33060
gtggggacga	acgcgggac	gtagcgttcg	ccgtgcgcgc	tcagggcac	actgcagccg	33120
tagagaagct	cggcgttcgg	gttgcggtcg	accgcacccg	ccaccgcgct	cggcgcgcca	33180
tcggccagcc	ggtcgccctc	ctccaccagc	acgacccatt	cgcccttcgc	gtccgcgac	33240
gcgcggttga	cgcgccccgag	cgccgcgtcg	ctcgcgctcg	gctcgaggac	cacggtctcc	33300
agggcgatcg	cggcagtgct	cctggccggc	aaggcctcgc	gcgccatcca	gtggtcgtag	33360
gcacgcgcgc	ccagcgcgct	cgcgctcgcg	cggcgtggaa	agccgaggca	cgccgcggcg	33420
aacaacggat	cgatcggttc	gggcacagca	tccaccgcgc	catcgcgctc	cgccgccagc	33480
tcgcgataga	gcgccaggta	ctcgcccgca	ttcgcttcga	tcgtcttcgg	gcggtgttac	33540
ttcaggttct	ccgcgacgcg	atcgagcgcc	gtgcggtcgt	cgcgcaggcg	cagcaggagg	33600
gccacgatcg	cgccgggac	gttggtcggc	agcaggaagc	cggtctcgcc	atcgcgacg	33660
cgctcgggaa	tgccgcacatc	gcgcgtgccc	gccaccggga	tgccgaggcg	ctgtgcctcg	33720
ctcaacacca	ggcagaaggt	ctcctccatc	gtcgagggca	gcaggaccag	gtcgatgcgc	33780
cgaaggatgc	gcggcaggtc	cgtcgcgcca	taggcgcgc	gctgcagat	gcccgcgcga	33840
ttcgccgccc	cttcgagcgc	gagatcgcg	ggaccccgag	cctcgatcgc	gatgcgctgc	33900
cccgccagcc	ggcgcgcgcc	ctcgatgatc	aggtgcgcgc	ccttgcgccc	gccgaagcgg	33960
ccgaggaagg	ccacgcgcag	caccggccgc	gcttcggggc	gcggaagcag	cggcaggctc	34020
gcaatgccgt	gcccgatgac	gcgcagcttc	gcttcgcagc	tcgcgcgcaa	gcccttgccg	34080
atgcgcacgc	ccgcgtactc	ggacggcgag	aggatcacgt	cgccgcgctc	cacggcatcg	34140
gacgcccgcg	cgaagcgcca	ttcgaggtag	ccggtcacgg	acgccggcgc	cgtcggcccc	34200
acgatcgctc	cgctgcggcc	gcgcaggcaa	tccacgcac	cgccatcggc	gcccgcgcgc	34260
gcggcgcgcg	cgcacggcag	gtcctcgcg	ccggtcatca	tggtgtagtc	ggcgcacagg	34320
tacgagaagt	cgctgcggct	gatgaccacg	cgtgcccccg	actcacgcgc	gatgcgcggc	34380
aggcgcgagg	tggtccagcc	caccaacgac	tggaagtga	cgacgtcgta	gccccgggac	34440
agcagttcgc	ggaaggccgc	ctccacgccc	gcgtcggcaa	ccgcgcgctc	gtgcccgatc	34500
acgcggatgc	cgggcgctcat	gcgctcgggg	ttcaggcgcg	cgagtcgcac	gtgcgtcgcc	34560
aggcgctcgg	tggtgaagtc	ggtccacgcg	cgcgggccct	cgggcagcac	cagcgtcgag	34620
gcgacgtcct	tgcgagcag	gttcaccagc	gcgcgcgtgt	gctcctcaat	gcccgcgcgc	34680
gagtcgaggc	gatgcagcac	ctgcagcacg	cgcgggcggc	ccggcatccg	ctcgcgctcg	34740
gcccgcggct	tgacgcgctc	ggcgggcgcg	cgcagcggtt	tcgcgatcgc	ccacgcgcgc	34800
acgcccgcga	cgtagggtgg	ccagcgcgcg	tcgaggcgca	cgcggttggc	ctcgcgcgcg	34860
gcgtcgatgc	cgccaccccc	gccgaacgag	gcctcgcccc	cgtggtgcac	gtagacgtcg	34920
tcgcagcacg	cgatctgcca	gtcccggctc	agcgcgcgca	tcgagaggtc	gttttcctcg	34980
ccgtagccgc	ggccatagag	cgagtcgaac	gggcccgcatt	gcacccacac	ctcgcgcgcg	35040
accagcatgc	agaagccgac	cgccgtcgcc	aggcgcggt	actcgcgcg	gctggcgcg	35100
gcgacgcagg	cggagatggc	atcgacgtcc	ggctcgccgt	cgggccgctc	gaagatatcg	35160
gcgacgccgg	gcaggctgag	gatcgctcg	cggttgagga	gcgggcacac	gatgccgacc	35220
cgcgcatcgc	tgctcgggca	acgcagcagg	ccctcgagcc	aaccggggcc	gacctgcgta	35280
tccgaattga	gcaccacgaa	gtcgccgccc	ctgcccggca	ggcccgcagc	gaccgcgcct	35340
gcgaagccga	ggttcttcgc	attggtgacg	acacgagtcg	cggcacgacg	cgccgcgaag	35400
ctgcccagca	acccggcgag	gcgcggatcg	gtgctcgcg	catcgacgag	gacgatgccg	35460
tgcgcgggcg	ccgtgtggcg	ggccagcgac	gcgaggcagc	gttcgggtgtc	gtcgtggccg	35520
ttgaagaccg	ggagcagcag	tagcgccggc	gccatccgac	cggcgcttcg	agctgggcca	35580
atcgcgccag	caggcgctgg	tgctcctcg	agagtgcttc	gagctgcttc	tgcgagcgga	35640
cgtaggcgcc	tccatgcgct	cccaggtgcg	tgcgcgcttc	ctcgagctgc	cgcagcgact	35700
gctcgtggcg	accggagacg	tcgtggctcg	ggtcgagag	cagcgagaag	cgcccgcccc	35760
ccgcccgcac	cgcctcgcg	tcgcccgcac	gggcgaggaa	gtacatcgca	tcggcgacgc	35820



34/39

ccgccgcggg	tttcgcagcc	tccggcgccg	aagcttgag	caactgcacg	gcgtcgggca	35880
ccgcgtgcag	cggccagatc	gccgaatacg	catcgacgcg	ttgccgaac	agggccacct	35940
ccgggaaccg	cgccttcagc	acggcgcgga	attcgtgctc	gtagagctcg	cgcacgtgat	36000
gcgggttgcg	gtagtcgcgc	tggtccgagt	acacctcgcg	gttcggcgtc	gagaccagca	36060
gcaggccgcc	gggcgccagc	acgcgcttcg	cttcgtcgag	caggcgctcg	ggatcgggaa	36120
tgtgctcgag	cgtctcgaag	gagacgagga	gatcgatgct	tgcgtcatcg	cacggcagcg	36180
actcgcagcg	gccctcgacg	tactgcaggt	tctgcgcggc	accatagcgg	cggcgcgcct	36240
gcgcgatggt	ctcggcgggc	acatccgcgc	ccaccacgga	cttcgcgcgc	gtcgccagca	36300
gcgccgagcc	gtagccctcc	ccgcaggcga	tgtcgaggac	gcggcagcct	ccggcgcgag	36360
gcaacgcaaa	gtggtagcgg	tgccagtgcg	cgtaccagat	ctcgccttcg	aagccgggct	36420
ggaaacgttc	gtgttccatc	gggagagtat	aggtggatgt	gcaaccgccc	tccggagagg	36480
gcggttgcga	tgatcttatt	cgagatcggc	tgctacggaa	gcaccggcaa	cggctgcggc	36540
ggcgccaccg	accatccggt	gcccgcgccc	agcaggttcg	cgctgccgag	cgtggtgagg	36600
ccgtccatca	ggcgcaccgt	gatgcggcca	tgcacattct	tgaacaccat	gtcgagcttg	36660
ccgtcgccat	cgaagtgcag	cagttgcgtg	acggtccagc	ccgccccgcg	cggcagcacg	36720
tccgccgcgc	tgaggatggc	ggtgccgttc	atcaggcgca	cgtgcgcacg	gccgtcggtg	36780
tggcggaaca	cgatgtcacc	gcgcccgtcg	cggttcatgt	cgcccacgtg	gctcacgata	36840
cagccggtgc	ccgcgcgcgag	gagctcgcaa	ccggcgccga	acgcggtgcc	gttcatgatg	36900
aagagatgcg	cgcgcccatc	ggtgtggcgg	aagaacatgt	ccgccttgcc	gtcgccgctc	36960
acgtcgccca	ggtgcgtcac	cgtccagccg	ctggcgggcg	agaggaagcc	cgcgcgcggc	37020
gtgatgggtg	tgccgttcat	caggtagatg	tagccgcggc	cgtcagcggt	gatgaagacg	37080
atgtcgtccc	tgccgtcgcc	gttgaggtcc	ccggtggcca	cgactctcca	gcccgtcgcc	37140
ggccccagca	gctgctggct	gccgatgatg	gccgtgccat	ccatcagcca	cagggtgcgag	37200
cggccgtcgg	cgttgccgag	caggagatcc	gccttgccgt	cgcggttcat	gtcggcggtg	37260
cggtcgatgc	tccagccgag	gccggcgccc	agcagctcct	tgccgcgggt	caccgtgagg	37320
ccgttcatcg	tgtacacgta	cacgcgcccc	tccgtgtggt	gggatccctc	agagtcgacc	37380
tgcaggcatg	caagcttgag	tattctatag	tctcacctaa	atagcttggc	gtaatcatgg	37440
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttcacacaac	atacgagccg	37500
gaagcat						37507

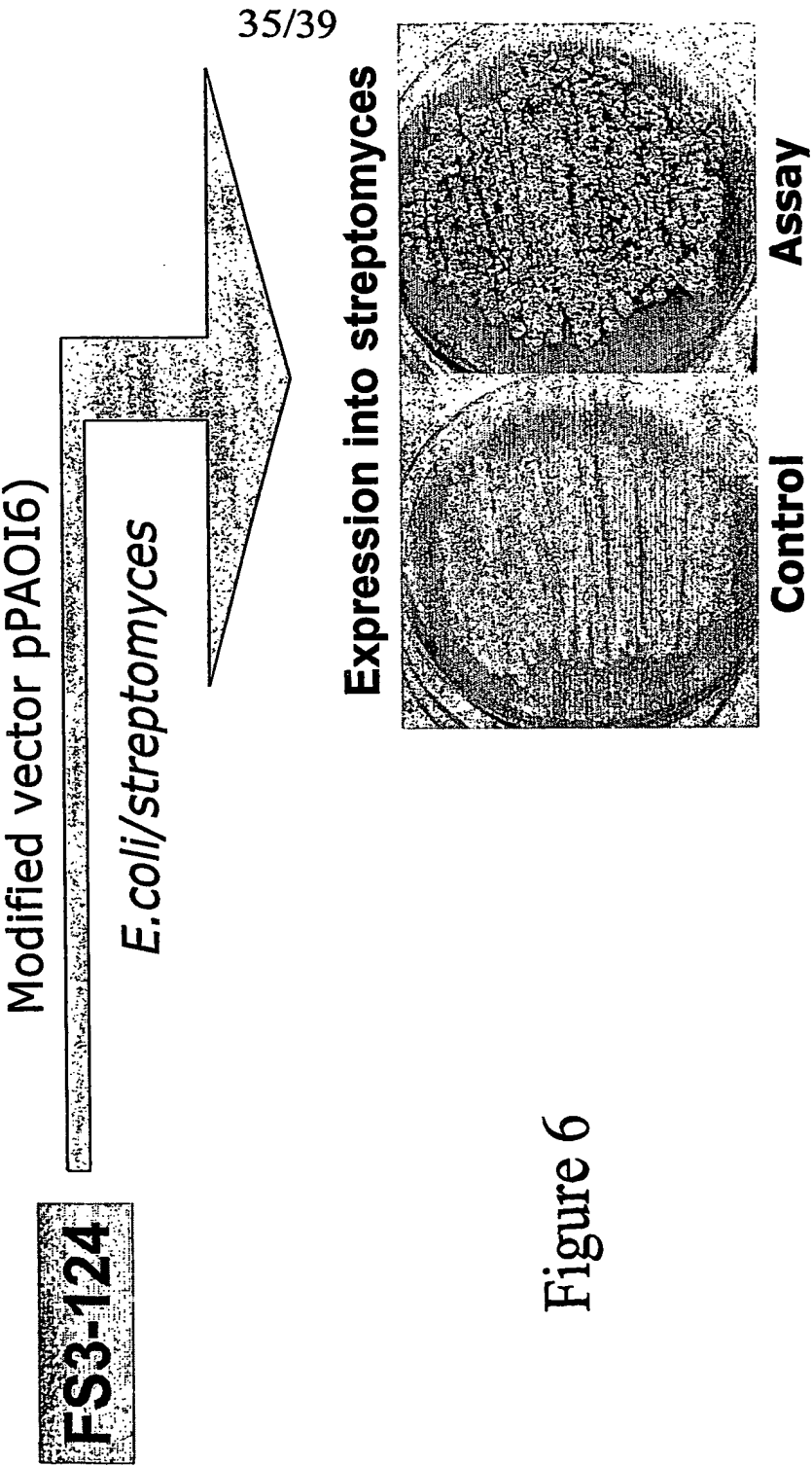
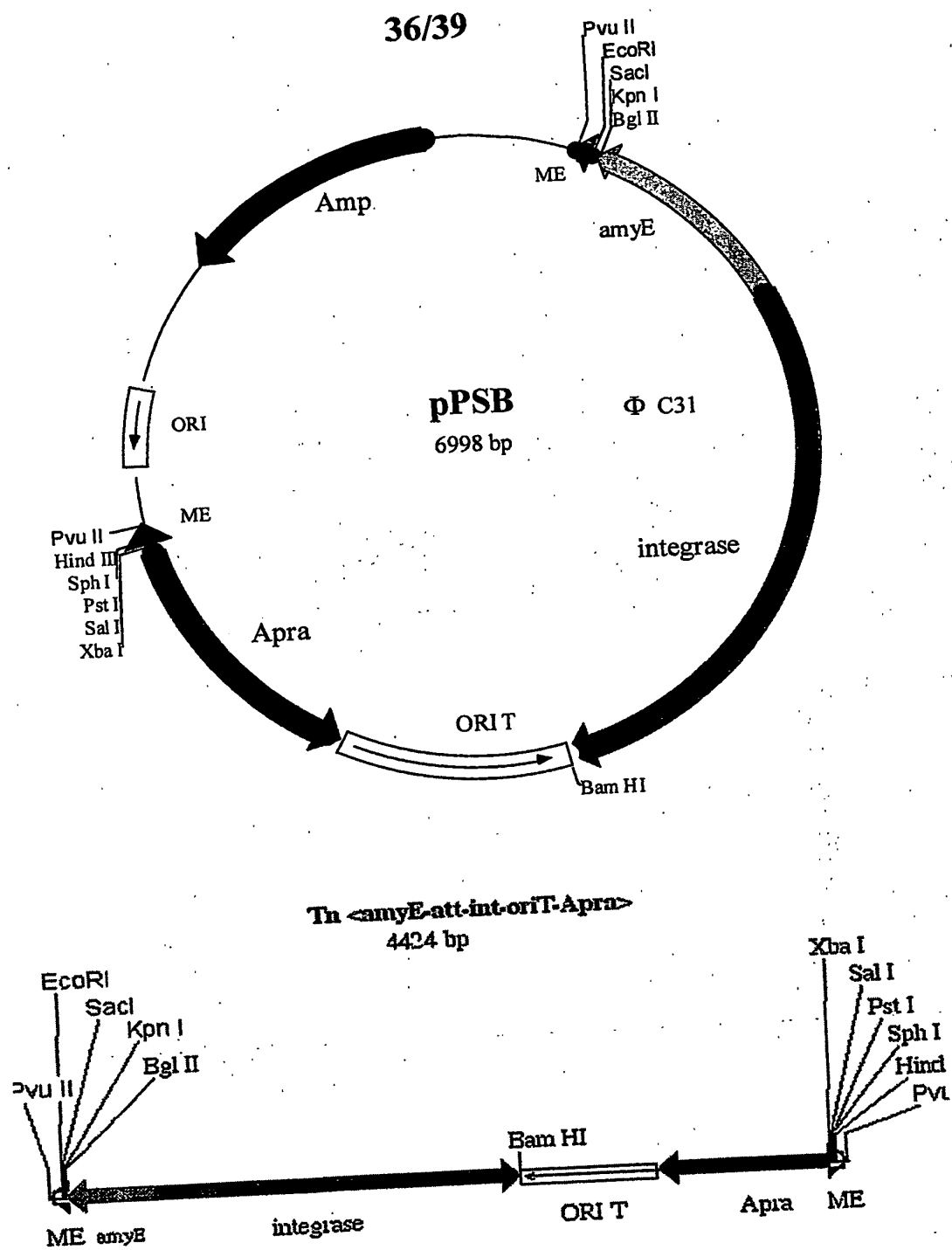


Figure 6

**Figure 7**

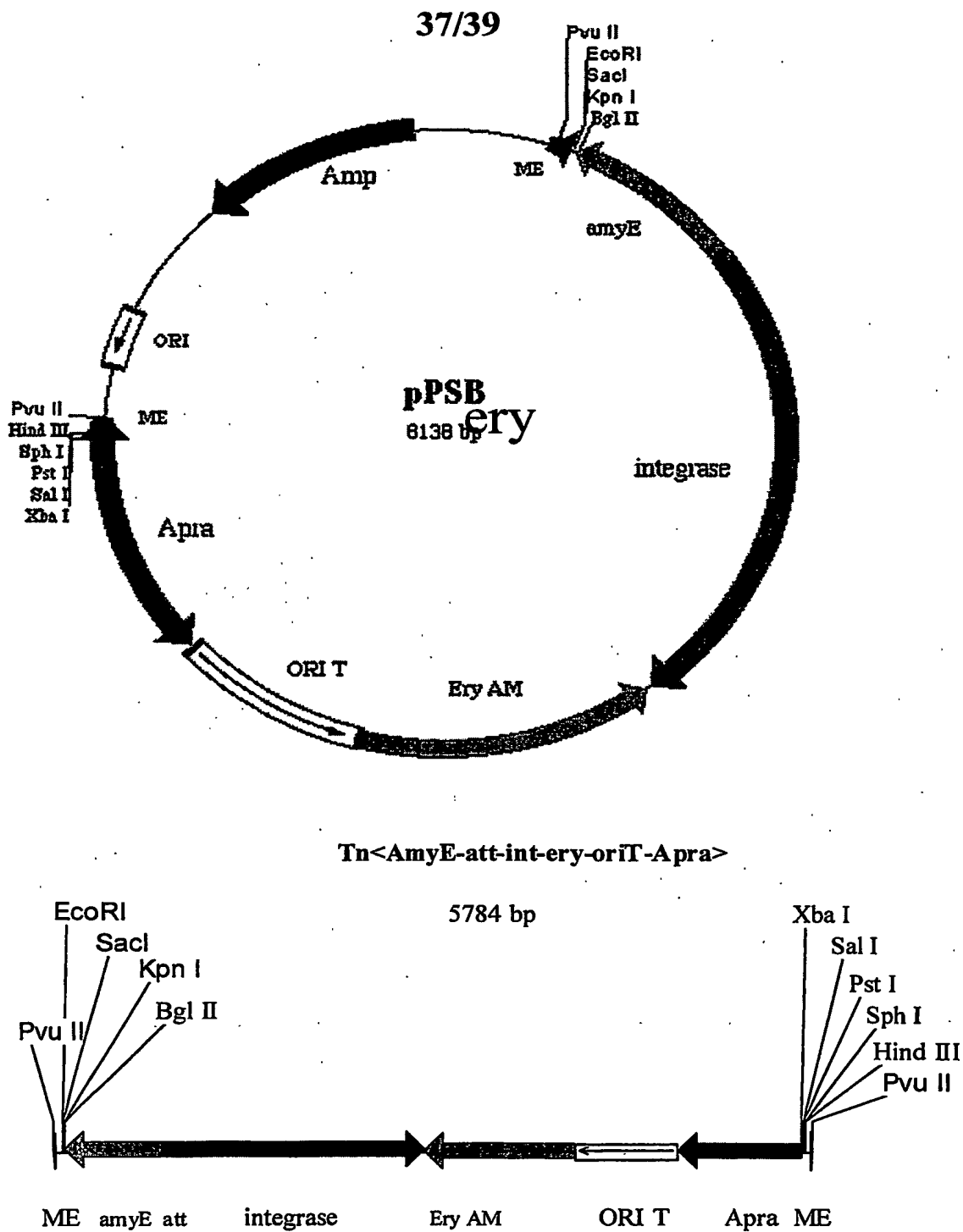
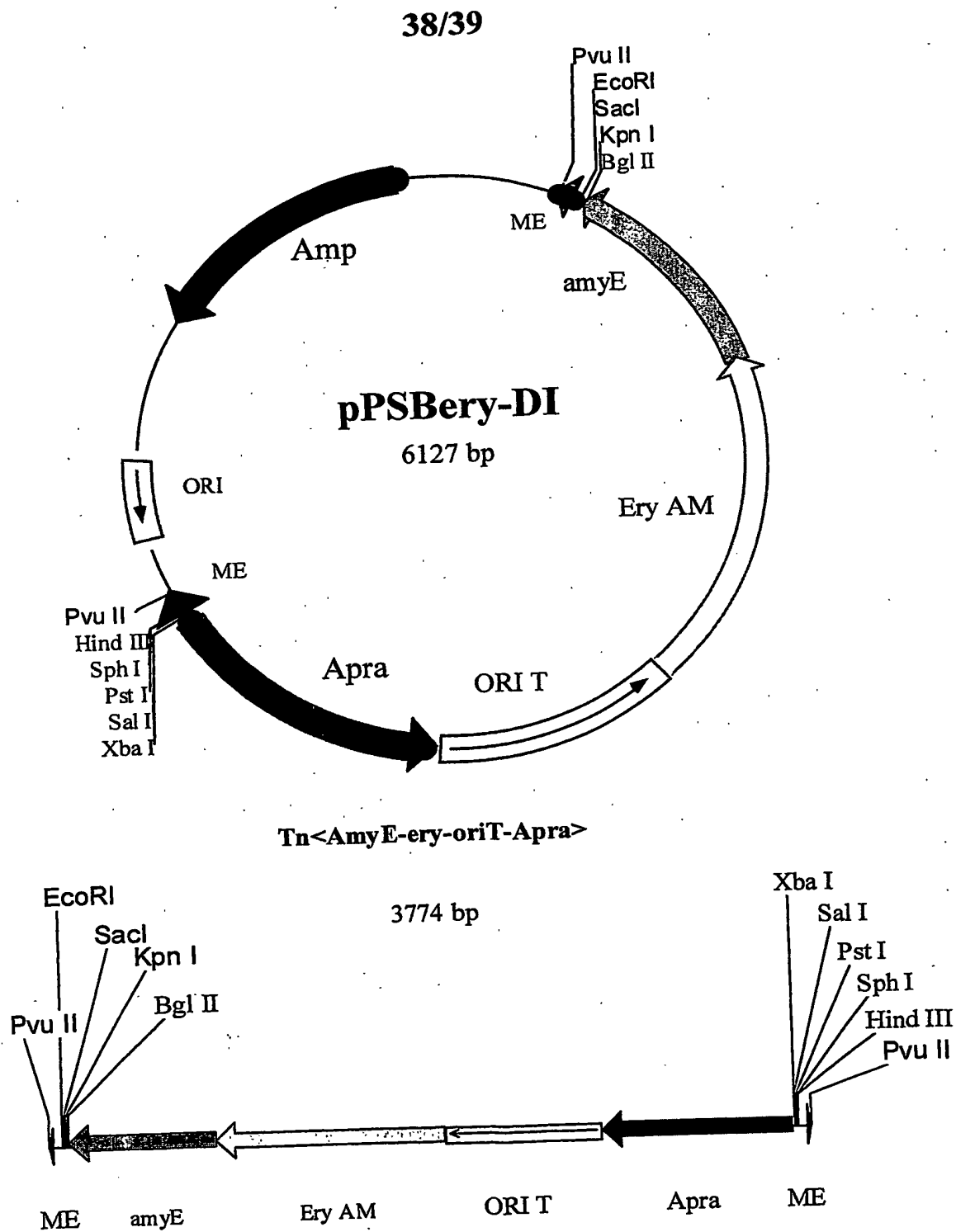
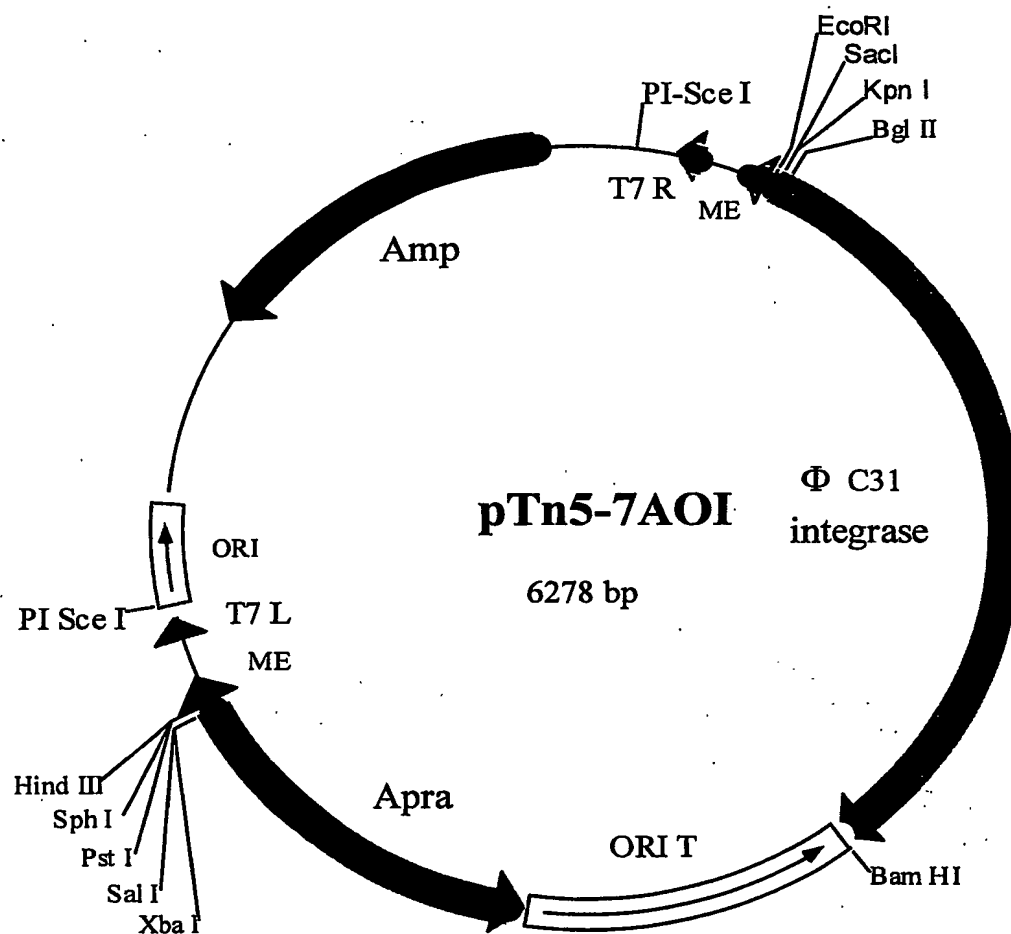


Figure 8

**Figure 9**

39/39

**Figure 10**

## SEQUENCE LISTING

&lt;110&gt; LIBRAGEN

<120> Method for the expression of unknown environmental DNA  
into adapted host cells.

&lt;130&gt; B0149WO

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 16

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 37500

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: DNA sequence  
of clone FS3-124.

&lt;400&gt; 1

gatcgggtgcg ggccctcttcg ctattacgcc agctggcgaa aggggggatgt gctgcaaggc 60  
gattaagttg ggtaacgccca ggggttttccc agtcacgacg ttgtaaaacg acggccagtg 120  
aattgtaata cgactcacta tagggcgaaat tcgagctcgg taccggggga tcccacgtac 180  
cacgagctca tctggaagag cccgctccag acaccgagg aatggaagcg cggcgctcgtg 240  
agaccgtaca cgcagaagcg tctcgtcgcc ttcgttgact ggcttgcgac gacgatgaag 300  
ctggatctca ccaggatgtt cgccgcaggc agctcgatgg gcggatcggg cgcgatcatg 360  
ctcgcgattc gctatcccgc acgattcgcg tggaccgtgt cgtgggtcgg cgtccacgtg 420  
cccgcgact ctccgcaatt cacatcgtcg tacgagctgg tgtaaggccg gcccgacttg 480  
aaggtgccgt tcgagaacgg cacgccggtc tgggatcatt ttagtgacgt ctggtacctg 540  
cggcagcatc cggagcagga catcgggttc atcacgttct cgaacggcaa gaacgactcg 600  
gcgatcgggt ggcgccaggc cgtcgaattc ctgaagacgc tgcaggagac caggcaaccg 660  
cacctgttcg tctggggcca ggaggacac ggccagcgcg cgaagatgcc ggaaggtggc 720  
ggcgaacggg agatgccgct cgacctcaga acgaaccaga gcctgccggc gttcagccgt 780  
tgcacgctcg acgacgatcc gggagacggg tcggattcga gcggcgctccc ggccgggagc 840  
atcaactcct acgtcacctg gcagcctgac tcgatcgtcg acgaaccgg ccgatggagc 900  
gtcgtcatca agctggcgaa tcgaccggcg cggtcgccc cgcagtcga cgtcacgccg 960  
cggcggtcgc agcagttcaa gctcaagccc ggcatcagg tgacatggac gaactccgcg 1020  
ggcaacaccg tcgtgcagcg cggcgaagcg gtgcgggacc ggtgggtct cgtcacgctc 1080  
ccgcagatgc aggtctcaac ggggtggaac aggatcctgg tgagcaggaa ctaggcgggc 1140  
ggcgctcctc acgagccggc gcatcggcgc cagttgatgt gcccacgaaa gctggctgag 1200  
gtccgcgcgc gcggcgatgt cccgtcgcgc ctgcggcggc tttgccagca aatccagcac 1260  
agccgccccg aacgcctcga ctgattccgc cacggagcac gccggccgtg cgatggtggg 1320

aagtccccct gcgacggccg gggtcacgac gttcgggagc cccgcggcaa tcgcttcgag 1380  
cactttgttc tgcacgcctc tcgcgatcgc aatcggcgcg accgcgatcg ccgatcgcca 1440  
gagatactcg cgtacctccg ggaccgcgcc ggtcacttca atcgtgggat cggccgcgag 1500  
tgcccgcacc gcggcggtcg ggtccatgcc cacgagcatc agccgcgcgt ccgggcgggc 1560  
gcgaaccacg gacggccaga cctctcgcgc gaaccagagc gccgcgcgct cgttgggctc 1620  
gtagccgaac acgccgcaga agacgaccg cgcttcggcg gccggtggte ctttcggcgc 1680  
aaacatctcg acgtcgatgc cgttcggcac gaccgacgcc gtgaggccgc ccgtgacttc 1740  
ttcgagcagc aggcgctcgc gctcgtcac gaccgtggte gccgcgcgc gccgcacgc 1800  
gactttctcg aagcggagca agcggccggc ctgcggcg aacagccagc ctttcggcca 1860  
cgccgcgctc gacgccagcg cgcgccattt ctcgagatcg acgtccacca tgtccaggac 1920  
gaaggggatc ccgccagga tcggctcgca ggcgtagcgg gccatccgc tcgagtacgc 1980  
gagcaccgcg tcgggcccgc tcgcggcaat ccgtcgttcg agcacctggt agatctccgg 2040  
cgaatgcagc aggacgtgcg tgaggggtcg atcgcccggc aggccagac cgctcgcgat 2100  
gagattccgg ccgcggcgga cgcgcacgac gtcggtcgac gccgtcacgc cggcgagatc 2160  
cgctcgatgc gacctcct cgtcgtcgtg cgcgagcgag acgaggtgca cgtcggcgga 2220  
tcgtgcgagc gtgtgaatga gatgaaacgc gcaatgcga tcgcccgggt tcggggcgta 2280  
gggcagccgg tgtgtgagca gaaggactct caccgtgcgg catccgatcg ctctcggatg 2340  
acgggggccc gtacgcccgc ggcgatcacg cgggcgggaa tcggcgcggt cacgaccgac 2400  
cccgcgcaa tcacggaatc cggaccgacg tcagccatga ccacggcggc gctgccgatc 2460  
caggtgccgg cgccgatccg caccatccgc ggtgtgcctg gctgctcacg gatcgccgc 2520  
gtgaggtcgc tcgtgccgtg cgtgtccggt ccgctgggga tgtggacgcc ggcgccgacc 2580  
aggacatcgc gctccagggtg gaccagccg agatggcatc cgggtccac gtacacatgc 2640  
tcgtcgagct tcgcgccggt ctgcgaaaag atggtgccga aggcgaccgt gaccgacggg 2700  
tcgcagtgtg caagcacgcg gccgaggaaa gccgcgcgaa ggtattgtcc gatgacgccg 2760  
gggatgagcg agagccactg cgtcgaccct tcgagggccc ggtcggcgcc gaggaagggc 2820  
cgccgcacgg agtaggagag caaggcgggc aggacgagca ccagcgcgat gccacgggcc 2880  
agtcccttgc cggcgtcctt gaggtcacg cccactgcag cgcgcgccc gcgtcggacc 2940  
gggctccgta tcgtgtgacg agcgagtcgt agagcgcttc cactttgcgc atccgcgcgt 3000  
cgaacgataa ctcttcttcg atccggcgtc gcgcgcaat cgcacgggtg cgtccggccg 3060  
cgggattggc catcgctcgc tcgatcgcgt ccgcgaggcc tgacgcccgg ccgggctgaa 3120  
tgaccagggc gtgctgcccg ttcgacacga gctcggccgt gccccggcc gcggtcgcga 3180  
cgatgggtgt ctcgaaacgc atggcttcga gcacggcggtt gggcgctcct tcgtagtccg 3240  
acgactgcac gagcaggtcc agcccgtgat ggacggccga gacgtcggac gtgtggccga 3300  
ggaaacgcca ggttcggggc ccgagctcgc gggcggcctg ttctcgcagc ggacgcggga 3360  
ggctgccgtc gccggcgatg acgagacgga ggcgcggcca cttcgcctga agggcggcac 3420  
aggcttcgat gagcaagtgc aatcgtttct gcggctcag acggccaacc gcgccgatga 3480  
cgaagtgcgt ggcctcgaat ccagtcctt cagcgccgc gccctcagat cgtggatcgc 3540  
gtcggaaagc atggtgggtc atgccattcg ggatcgtgac gaccgatcc ggccgtccgc 3600  
ctttggccag gagctccgc cgaatctgtt cggacacggc aatgacggca gggaagcggg 3660  
cgagaatccg gcgatcgagc gggtagtaga cgggtcgctt gcgaaacgaa tggccggtcc 3720  
atccatgggc cgtggccagc ggaatcacgc gctcggcgcg tcccaggagc agggcgagga 3780  
gatcgggtct gtactcgtgc gcgtgcacga tgtcgatgcg cagatctttc acgagcgcgc 3840  
ggagcttcga ccagattccg gtgtcgaacg aatgcttctc ccggacctcg acgtagtcca 3900  
tgccggccgc tttcgcccgt tcgtccatgc cgaacaccgg atcgcgctgg tcgcggatgt 3960  
agcagacggt gatcgcgta tggagcgat cggcctgcgc cgttccgagc agaatcgtct 4020  
tctccggtcc gccgccggtg cccgcacgc tcgggagctc gagcacgcgc acgggacgcg 4080  
cgattcccgg cctaggcgcg gcgtctggca gcatggcccc cccggccgac gagcgcgatg 4140  
gcccgggcca tatcgtacgt gccggccacg acgagcgact gtgcgatctc gacgagtccc 4200



actgcgcccgc gccggcgccc gatcatcgcg gctgcgcgca agccggcgaa accggcaagc 4260  
agcaggagcg cggtagcgac aatccaccac ccgccgcgct gccacagcgc gaggcccgagc 4320  
ggaatcgaga cgagcgccag cagatcgagg accgggatga tcacgctggg cagatcccgc 4380  
atggcgagcc ggtgacggaa gctcacgcgc aggttgctgc gcccgcgcca cagctcgccg 4440  
aaaaacacgg cgcaagcga tcggggatcg ccctggtgca cgctgcggag ccgcgggtcc 4500  
gacaggatgc ggccgcccgc ccggcgcgag cgctggcaga tatccacgtc ttacacaggtc 4560  
tcgaggctgg tatcgaagcc accggcgcgg tcgaaggcct cgcgcttcac ggcaagattg 4620  
ccgctcggga gccagtccac gtcttcggtg ctctgccac gacggcgaaa cgagtcgtac 4680  
gcgcgctgca ccagttgcc gtcaccggga gcttcgtact gcgcgcggc agcgacgacc 4740  
cccggggtct ggagcagatc gacggcgacc accggccagc gcgggtcgat ctcatgatcg 4800  
gcgtcaacga acgccagcac gtcgccctcc gccgaccggg ccgcgcgggt gcgaagcgctc 4860  
gccacgttca cgtcgggcag cgacagcact cgcgcgatc cgcgctctgc cacggcacgc 4920  
gagtcgtcgg tcgaccggtt gtccgccacg acgatctcgt accagccagc cggcgccctc 4980  
gtgcgatgaa tcgacgccag gcaacggctg aggtgcgcgc cgccgttctt gacggggatg 5040  
acgaacgaca cgcgcgcgcc caccggcaatt ttccgggtca tcgatttcgt agcgcgcccg 5100  
gctttcgtag ccggcgcccg gttgggggcc gcgagaactc gaacgcgccc agatccggcg 5160  
cccctgctct ccggcgccga tcgaaatcga gcggcgccct cggaatctcg atgccggcgt 5220  
cgatcgccctc cctggcggtt gccgacaggt gcagatcgca cgaggcccg tttgcgaacc 5280  
actctgcggc ggcgttcgtg acgttggtcg tcgcctccgc ctcaccgccc ttacggcgca 5340  
ggatctgcgc gctcgtcagg ttgttgcgga tgagcccggt cgtcacggga aaccggacgc 5400  
cgatcgtcca cggcgtcgcc gtgccgagcg tgaagacgct gttgtgctcg atccggaagt 5460  
cgggtggacgc gttggcttcg atcgcttcgt ccggccattc gttcagggtg cagatcacgt 5520  
tggtccgcac gatgccgcgc tggtagatga tctgcttctc gccgtcgcgt gcgtaggccg 5580  
tgatccccgc ggccaggccg atggcgatgc ccctgaacga atcgacgatg acgttccggt 5640  
cgatgagcgt gtcctgcgag ttccgccagg cgaggatcgc cgggcccgat cggcagccgc 5700  
ccgattgcgg gccgcgaatc cgcaggaaga cgttgctcgc gatgaccac ccccgggccg 5760  
ccaggatgtc caccgcgtcg gtgtaatcgg accgcgcgct cgtcgtgtac tcgaagcgcg 5820  
agcacgccac gaggccatcg tccgcgaagc gtccgtcagt gccgacgctg cctttgagca 5880  
attgctgtcc cgcacgatg agctggacgt tgtgcaccgt ggctcgggac gcgccaagct 5940  
cgccgcgcac ctgaatcgcg tggtagccga cgtggccgac cgtcagatcc gcaatcgtca 6000  
cgtcagtggc gccaaaccgag agcgccacgc cgatcgctc gccgggtcat ccgcgcgcgc 6060  
gaatgacgac gcggtcccga tcgccggagc ggctccggag cacgggtgcg gggacgttga 6120  
tgctggccat ccggtcgagc cggtagtcgc cgttctcgag cagaatcgtg gtgcgcggct 6180  
tgacgcgttc gagcgccctg agcagctcgg acgtccgcgc caccggtgacg acctgttctg 6240  
tcgccggcgc cggccatccg cagaagtgcg gcaccggcgc ggccggcctg ccggacggca 6300  
gaagcgccac gagtgcgcgc gccacgatga ggcgagtcac gccgccaccg tgtcgacgtc 6360  
gtgttcgcgc aaccagagct cgaggatgag caggatttga aggaggtagc cgcgcgactg 6420  
tctcccgtag ccgctgtcgt cgatgatggc gcgcaacgtc tcggggcgca ggaacatccg 6480  
cgtgcgcgcc tcggggcgca acagctcgc ctcgacggct tcgagcagtc gcgtctgcag 6540  
ccagcgatcg aagtcgtgat agtgacggtc cccaggcgcg ttccaggcgt tcagcagggt 6600  
gttgaccttg tcgaacacct tctccgccag cggaccggcg tcgccgcgcg ccgggtgtt 6660  
cgagttccgc accttgagca gcgcggggtt gccggcctgc gtgatcgcg gatgaagcgc 6720  
cgtgtcgtcg cggcatgcc gatcgccgct cagcaggacc tgcaggaaact cgcgatccat 6780  
gaacggcagc cgcacgtcca cgacggtgcg gaagagatcg agcgaggga tcgtgaagcg 6840  
ccgatgatgc tcacgaagat agagatagct gctgagctcc gcgggtgaga gctccacggg 6900  
gtcgaggagc gcgtgcagc agtcggccgc gcccgtccg gagaggcgcg cagcgtctgc 6960  
tgtgaagagc gtgtccagc gcaggccgct gctgatgtag ttggcgcgcg acgtcatgta 7020  
gtcgatgaac tcgccgcggc ccgtcatcgc gaacacgtgg ccgtcgggtg ggagcggcca 7080

cgccaggetc gtcttggcga gctctccgcc gtggccgcgc agcaggacct cgataccgga 7140  
gtcccgcagg aaatcgatcg ccagcatctc ggtcagtcgc tggctcagat acatgccatc 7200  
ggtgagcgag accatgcgcg cctgggttcgg cagaaagtcc ttgaggtacc ggtcgtcgag 7260  
ctcgaagaag cgggtggtcgg taccggtgag cttcgacagc tgttgcgcgga tcacctgac 7320  
ggcgacgccc gcgacgccc gctgtgacgt gatgagccc ccaacaccgt tcaccgcaact 7380  
cagcacggcg cggctgtcga ggccgcccga gagcgacagc ccgaagcgat gcctgcccga 7440  
gagcgacttc tcgaccgcgc gcctgaacgt cgtgtgcacc gcgtggacgt acggtcctt 7500  
cgctgtacgc ggtccgcggg agagatcggc cggctcgaaa taccgggaca cgctcgcggc 7560  
gcctgacggc cagtcgtacg cgagcacggc gccggggctc aggagcgtga cggccgccc 7620  
cagggctctt tcgcccagga cgaagcccag cgtgacgtaa tcggcgggccg cccggaggct 7680  
gagagccggg ttgcgcgcgc aggtcctgag cacggaagcc agatccgac cgaacgacag 7740  
tccggccgcg tcgaccgcgc agtagatcgg atacgagccg aacggatccg ttgcaagcag 7800  
cactcgtctc cgcgcgggat cgatcacggc gagcgaaaac gcgcggtcga gccgggccc 7860  
catggcaggc cccagccgtt cgtacaccaa agccaggaga tcggcgcgcg tcccgttcgc 7920  
cgaacaaaac tcgctccgga gcgcgcgttc gttgtagaga tcgcccgcaa agatcaccag 7980  
cgcccggtc acgtcgcgcg cgccgcccac ctgctcgagg aagccgagcg tgccgagccc 8040  
gaccgcccac cggtcgcagc gatctcgata gacgaccgtg cggcgcgggc tgcccgactt 8100  
cgccgctgac gcggcgctcg cgcggtgaa ggcggtagat gccggctcat cgcggcaaac 8160  
gacgcccagg aggcggttca tcagacgacc ccgcgggcag acacgcgcga cggaacaaga 8220  
ggtgtcgtct gagcggccgg aaccggccgc gcggcggtgc ccgccaccat ccggttgagc 8280  
gcaatcgaca agccgaggaa gtgccagagg atctcggtgt actgaaagat gacgaacgtt 8340  
ccgcccacgc agaatgcaat gagggcggtc tggacggcca gcgcgtacgg cgccagctcg 8400  
gccagggccg gatcccgtt cgagaccggc cggaccctgc ggcaggcgat gagcgccagg 8460  
aggagagtga gcaggaaacgc cgcgaggccc acgaacccaa gctcagccac gagaccgaac 8520  
cacgcgctgt gaaccgagcg ccccttgccc caggcgccgt ccgacgtgtc gtagtcgttg 8580  
taggcggccg tgaaggcatt gtgaccgacg ccgtaaactc gccggtcatc ggccatgcgg 8640  
agcgcgaccc tcagaagtgc cagccggctc tgggccgacg gctgatcgcc aaccgcccgt 8700  
tcgctcgcgt cgatgctcga gtcggacgag ctgatcgagc tcatgcatc ccagaactct 8760  
tccgtcatca cgagcgacat cgcgacgaag accgcccgcg cggccacgag gacgagcatc 8820  
ttgcgcttcg agtacaggaa atagacgagg ccaagcacgg cgagcgtgag gaagccgccc 8880  
cgggagtacg tgctgatgcc gcgatagagc acccgatga gcagcacgcc cagcgcccac 8940  
ttggcccatt tcctcgcttc cgtctgcac agcgtgacga tgagcgccac gagcatgaac 9000  
ccgggaatga acaggagctg accccagccc tgcttgccgg cctcgaacc gagggagagc 9120  
gagatcacga ggaagacgag ccgcaaccgc ttgacgtccg tcgtgaggac tgcgagcagg 9180  
tacgtcatca cgggtgactt catgaagtcg atcatgtagc ccacgcgaa gtcctggtag 9240  
ggcgacgcca gcgtcgacga caggctgtgg agcccgaaga aggcgaggag cgccagccgc 9300  
gcgtcgatcc gcagcgctg cccggacagg aacgcgatcg cgagcgtgta catgcccgc 9360  
aggagcgaga tgttcatgtg gatgagaaag tcgctccaca ccagagctc cggccggaag 9420  
tacgcgatga acaggtagaa cagaatcgcg tagaacggcc cgcgcagcgc gtggaaggcg 9480  
ccgaaggcga ggagggcgag gacgaaagcg gctcgaagca ttgccgtgtg ttacctgtta 9540  
tcgggtcccga tgggtccagag cggctgcgtc cgggtctcct ctacattgaa cacctttag 9600  
agggccggaa tcgaggtgaa catgagcacc acgaaaagaa tcgcccagac caccatgtag 9660  
gcgaagaacc cccgctgctt gtagagcttc tccggttct gcacggggct gttcggcagc 9720  
atcccgatgt gcaggtagta ggcaacatg cccgcccga cggcgcgaa caggatcagc 9780  
tcgaggtgat agcgagcat gaagatgcc ccgaagagg cgacgaggt ggcgtagaac 9840  
accatgctca cgagcagccg ctctcggtt tagaaggcaa acgacttccg gtacgaggcc 9900  
gcgaccgacg cgttgccgat gtgccggaac tccggttacc gttcgtcgc catgaagaac 9960

gcgcccacca tccagtagca gatcacgagc gacatcggcg gcacgcgac ctagaatcagc 10020  
gggaaccagc cgaggaggag ccggacggcg ttgttcgccc actcgctcag cacgtcgagg 10080  
tacggccact ccttcgtgcg aatgggcggc acgttgtagc tgacgccgag caccagagc 10140  
gccagggccg acagcgcgaa gtatctgttc acgagcagcg cgagcacgaa tccggccacg 10200  
cccacgagga tccactcggt gtagcccgcc gccggcttga tcttcccgga gggcaccggc 10260  
cggtgcccgtt tctccggatg cagcagatct ctcgggcccgt ccaggagctc gttcagcacg 10320  
tagttgctcg aggcgatgag gcaggtcgcg gcgagcgcg gcgagcgcg cgggacggcg 10380  
gtccagccga agagctcgcg ttcgtagaag aacgccagca gcacgccgag cagcatgaac 10440  
gcgttcttga accagtgatc gatccgcgcg atctggacgt acggccaaat ccgcgaaccg 10500  
gagctagcgg ctgacgacat cgttgagctc ctcgagccgc cggatcgcg accatccgcg 10560  
gccggctgtg acgccgtggc ggccggcgat gaggggcgag gagagtcctg cgggcatcgc 10620  
tccttctccg tccacgtccg gtgatcccc gacatacagg acctcccag gtccgagtga 10680  
ccacatctca cacgccacat ggaagcccg cggtgcccgt ttcagggcgt tcacgttgct 10740  
ggcggtcgtg cagagcgcca gcgaaaaatg ctgcgacgtg ccgagcgcg cgagcttgct 10800  
ctcgggcgcg tagtccgaca gcacgccgag cctgagcccc gccgcgcgaa gcgagtcgag 10860  
cgcgccgagc aaccccgggc gccggcacca ccgcagatac ttgagcggcc ggcgcacat 10920  
ccattcgttc accgtcgccg ccacgacctc gttgtccagc ccgagcgctt gcgcgggtgcg 10980  
cgcatcttg cgcgccgcga gcggctcgtc ggccgcgccc aggcgcggga gatcctcatg 11040  
cgccccggcg tactcccga cgatgcgcgc cgtgtccact ccgcgccgc accggaacgt 11100  
cacgagcgcg gtggcgcgga gctccgctgc catggcgcg cgcaacggac cctgccggta 11160  
cagcgtgccg tccacgtcga acagcacggc cttgacccc cgcaacggct tcaactctgcg 11220  
agggccagca tccgcagaat ccgcgacacc tttgacgcgt cgcaaggaa cacgtcacga 11280  
accacacagg cagccagccc cgccaacagc accgccgctg ccgcgccgc tgcgctccgc 11340  
cagctcccgc gccaggcgcg cgtgtcgatc gcggcgcgcg cgtagcagag cagaatcggg 11400  
acgagcgga ggtgataccg gggatgtccg aagacgatc cgtgaatgcc gctcatgaac 11460  
acgatcagcg tcaggagcag caggtgcgcg gccagttga ccggccgcgc gaacgagatg 11520  
ccgatgacgc cgagcaccat gaccgccgcg tacgatccca tgatggacac cgcgccggcg 11580  
aagaacagcc accgcggcg acggtagaac ccatttccga cgccggcgat gaagtgcgc 11640  
tccagcccc agaagtctgc gaacttcagc accgaacggc gaagcgtcgt tcccggatgg 11700  
gccttcatgt agtcgagcgc ctggcgtgc gccacttct ctttcgtgc ctccgtccag 11760  
cggtgcccgt ccggcgcgcg aggcggcatc gtgcgcgacc attccttgct gcccgtagc 11820  
gagatcgct cccacatgcg atcctcgggc gtgtacgcgt agttgcccac catcaggttc 11880  
agcccgcga gcgtgtccac gaccgtgaac gagcgttgca ggagcgtgtt gcgcacgctc 11940  
cacgggcca cgatcgcggc gtatccggcg aagagaagca cggccacgcg cagacgcgcg 12000  
cgaagcgga tgccagacc gacgagcgcc gcggccatga ggatcgcgac gtacggccac 12060  
atgatcgagc gaacgagcgc ggacgcgcg agcgcgacgc ccgtcgcgag cgcatccac 12120  
gccgagtgc gccctgatgg gcgatcgat agcgtcagac atcccaagag caacaaaaga 12180  
aggagtgtga tgaagagcgt ctcggacagc accagcacgc cggagaacag cagcgaggga 12240  
tagaacgcga acccgccgc ggcaatgagg ccggcgcggt cgctgaacaa cctgcgccc 12300  
atgagataca cgagccagac accgagcagg ctcaacggaa cctgcgccag ccggaccgcc 12360  
gtcagcgatt ccgagcccgt gaccgcccag acccccgcga cgaaccggg aaagagcggc 12420  
gcccggatgg acgtcggttc ccccgcccc caccggaacc cgttcccggc gacgacgttt 12480  
ctggcaagcg tcgcgtagt ctgctcgtc cgaatctcga gcgcgacgct cttgaacccc 12540  
caaatacagg cgagccggag gaccagggc agcgtgagga tgacgagaat tgcccggcgc 12600  
tcccaatggg tgcgctcggc ctcgctcgc ccggggcgga cgattggcca gggggccgac 12660  
gatgggagt ggggagcct catcattcgc taaaacaagc agttacggtc aaacctgtgc 12720  
caccggggaa tcagcaaaga ctctgccaac gggtcgccg gcggcgacgc ccggcgcgcg 12780  
ttttccggcc ggaagtgcgc ccccaaggcg acccggtg gcacagccca gtagaaatct 12840

cattctggga tgacttttgt ccatgatoca tgaccatttc gaggggcctt cgtgacgccc 12900  
tggggatggc ataagcttgg cacctgcat ggagacgtgt cgatgatcag cggccgtgag 12960  
gagctccaga aggcctatcg ggacgaccgg gtcgcccgtg aatacgtggc gcgccgggttc 13020  
cagtcaccgc tgggggcgct cctccattcg cgtcagatcg gggtcgtgog agagctcgtc 13080  
cgcgcgcagg gcatccggcg cgcggcggag atcgaccccg gtccagcgcg gctgaccgtg 13140  
gacatcgcg cgctcctcga ccgctgacg ctctcgacg cgagcgccca gatgcttcac 13200  
gaagcgcgog agcggtcgog cgaccggacg ctccgcccgc gggccaatct cgtgcaggcc 13260  
gacgcgtttc gtctgcccgt cagcgggcog ttgatctcg tctacacctt caggctcgtg 13320  
cgccatttcg agcgcgccga tgggctgcca ctctaccgcc agattgacgac gattctcgga 13380  
cccggcggtt ggctcgtctt cgatgcggtg aatgagattg tgtctgcgcc gcttcgcgcc 13440  
aacgcgaagc ccggcgagta cgagcactac gatgcgctgc tcacaccggc cgggctacgc 13500  
gacgagctgg ccgaggccgg gtccacgatg gtgtcgctca ccggtgcgca tgcgccgttc 13560  
cccacgctca tgaagtcca gatgtatctc gcgccgaggt caacggtgct cgcgcgcgcg 13620  
gcgatggagg tgatcgatcg gctcggcggc gaaccgctcg agtggatcgt cgtatgccgc 13680  
cgcgggtgac ctactggacg ggtacctggg atccgatcaa agaagcgatt tcgaagggaag 13740  
tgaacgccct ccgacccggg tcgcggcgca acgogacggt cgtgtcgttc tcgccggccc 13800  
agcgagccg tctcgtgccg aaagatcgcg tgctcgtgtt gtccggccgc gcctggatcg 13860  
cccttcgggc cgcggccgcc gtgctcgagc cgcgcggcga cgtgacgcac gtcttcggcg 13920  
gcggtttctc ctggcatctc ctgagggcgc tggggcggcg gccggttctg ctccaccgcg 13980  
tcaaggaaga tggcgtgccc gatctgcccg gggcgatcgc gcacgtggcc gtcgaagtcg 14040  
gccgcgctcg cgacacgtgg atcgccggcg gtatgtctcc cggccgcgtg cacctattc 14100  
atcccgccgt ggacctcgac tggatccggc cgggtgccgtt ctctcgag agattcacgc 14160  
tgctcttcgc cagcacgcgc tccgatccgg ctgagatcga gccgcgcggc attccgctcg 14220  
tcgtcgagct ggcgccggtc cgtccggaga tcgacgtgct ggtgccgtgg cgcgggtggg 14280  
gcgacgtcga cgcctccagg cgcgcgctcg acgccctccg tccgccaccg aacttcacg 14340  
tctcgtacga ggacgcaccc gacatgcgcg cctgttacgc gcgggcgcac gccaccgtcg 14400  
tgtgtctcgc gggcgccacc ggcaaggcct gtccgaattt cgtgctcgag ggattggcca 14460  
tgggacggcc gtgtctctcg acgcgcgaga acggcatcgc ggagatcctc gatcgatcgg 14520  
gcgccggcat catcgccgcg cgcgatgccg cgtcgctcgc cgcgagcctg gacacgctcc 14580  
gcgccggctg ggccggctgc gccgagcgcg cgcgccggct cgcagaggag acgttcgacc 14640  
tcggccgctt ccgcgccgcg tacgacgcgc tgtacgtgga ggtggcatcc gccgccgctg 14700  
gcgagcacgc gaggcgccgc tcttgacgaa gggactcgcc ggacggctgc tcgggtgggc 14760  
tcagaaggcg aacgatttcg ccgactacgt ccgccatcga tcgagatga agcgcgacgc 14820  
gcgggcgaat cgcgcgaggg tccatcgggc gctcggcgcc cgcgtgaacc ggatcctcat 14880  
cgtgcggctg gccagcatcg gcgacgtcgc gcgcgcgacg ggcgtcctcg gggccctcg 14940  
cgcgcagtac ccgcaggcga cgatcgactt cctgacgtcg gacgcgaccg tggccgtcat 15000  
ccggcaccac ccgcgatca acaccgtcta cagctccag gatctgtcgc gcctgtccga 15060  
gtacgactgg atcatcaacc tccagaacct cgtgccggcg gcgagctttc tcagaacgtc 15120  
cgggctctcg taccgcgaga ttctcgagca gctctcgagc cgcagccgtt ccaggatcat 15180  
caccggccgt cacctcgaga acggctcgga gtccacgccg acgaacatcc tctactgcat 15240  
ctgtgagatc gaggagcatt tctcacggc gctcgtcccg tacgacacgg ccgctaccc 15300  
ggtgaccgcg attcacgtgg acgcacaggc acgagcgcg gcacgatcga agttcgcgtt 15360  
cccggagAAC cgcaggtgc tggcgatctt cctcgggtcc aactcgtcg ggtgcggcg 15420  
cgatgagggc tttcgacgt actccatcga ctacgtgaag cgcctcgtgc gtcacttcga 15480  
agatcggttt gcgatcgcca tcatcgggca gtcccagggt cgcaaccgg ccgaacgcgc 15540  
acagtaccgc gagatgctcg gccgccatcc ggacgtgac gatctcgtgg acaagacgtc 15600  
gctcgacgag ctctgggcc tcatggactc gttttccctg ctcatcagtt gcgactcgag 15660  
ccccgtccat ctggcgatgg ccgcggagt gccggtcatc gggctgtacg tcaacgacgg 15720

gacgttcagg atgcacccga cgctcggcga cgagcgcttc gtggccatca acagcacgcc 15780  
gccgtgcttc acctactcgt ggcgatggaa gttcttctgc tcgacgtgcc gcgacccggc 15840  
gaccagatcg cattactgtc acaacgaggc gttcgtgttt ggcgtggacc gcattccgct 15900  
ggccgcaatc gaccgcgcgg cggcgcgggt gctcaagggg agctccgatg caggacctca 15960  
ttcgcgcgca gtttgaagag tcgatcgaga cgaagcgcaa agcgctctcg ctctcgatg 16020  
cgggtggcga gtcggcgcg ctcctcatcg aggcgctcga ggccggacac aagatcctcg 16080  
tgtgcggcaa cggcggtcgc gccgcgact cgcagcattt cgccgccgag ctctgtggcc 16140  
gcttcgagaa ggagcgccgc gcgttgctcg ccgtggcgct caccgaccgac acgtcgatcc 16200  
tcaccgcgtg gccgaacgac tacgcgtacg agaccgtgtc agccggcagg tgagtgcctc 16260  
ggagcgccgg gcgacgtgct catcggcac acgacgtccg gcaactcgcc caacatcctc 16320  
gccgccatgg acgaggcgga aaagaaacgc atgaaacgaa tcgcgctcac cggacgcggc 16380  
ggcggcaaga tgaagagtct cgccggcgctc tgccagggtc tcgtcccgtc agatcagacg 16440  
tcgaggatcc aggaagtaca cattaccgtc atccacatct gggcgaagtt gatcgagaaa 16500  
gcgctcagct gagctggcgg tagacgtcga acgtgcgatc ttcgaacgtc gaccaggagt 16560  
gctcgcgggc gatgaggctg gcggcgcggt cggtgaggcg gcggcgagg tgctgacctc 16620  
cgaacaacag ttgcagtoga ttgcgtagc cctcggcgctc gtccgcctgg acgagaaagc 16680  
cgggtcttccc gtcttccacg acttcacgca cgcccggtat ggtggagccg atcacgggca 16740  
cgcgggccga ctgggcttcg agcaccgaga gcggcagccc ctctgtcggtt gacggcagca 16800  
ggaaggcgtc ggcgccctgg agcagcaccg gcgcgtcctg gcggaagccg agaaagcgca 16860  
cgcgatcctc gagcgcgagc tcggcagaga ggcgctcag cgctgtctcg tgttctctgc 16920  
cctcgggtccg gtcgctcccg accagccagc actcgaccgg caggccgcgc gacatcagca 16980  
gcctggtcgc ccgattgccc gtagcctgcc ctttgtgtgg cgcaagattc gcgagcatga 17040  
ggacgacgaa gcggtcggtc tcgaggccgg cgcccgtagc cgcggttgc cggtcgccgg 17100  
ccacgaactt ctcccgatcg acccggttcg gaatcgcggt gaccggcacc ttgtcccggc 17160  
gtccctgctg cgcccgctcg atcatcacgg cgatgtagcg cgcacacgtg acgatgtgct 17220  
gcggcgggct cttcaggacc caggcgatct cctcgggata cggttcgatg tgaaagtga 17280  
cgccacgcgc cgcgcgcgcg gcctgcagcg cgggtctcat gaatccgtac acgacaggat 17340  
tgtgcacgtg gaccaggggc cgcccgacga gaccatcc cggaagcatg cgcgcacacg 17400  
cgcgagctg ttgcaggccg ccggcccga gcggttcag gtcatacatc cgccgccgga 17460  
cgccctccgc gtcgagcgcg gcgcccgcag gaccctcgcc tgtaatccat gccgtgcacg 17520  
ggacgcgcgc gcgtttggcg gcggccgaca accggatcgc cagcaccgcc gcgcccga 17580  
ccagatgtga gacgaggatc tcgtgcaggc ggggggactg attcatggtg gcgcgcgatg 17640  
cgcccgggcc gccagcctac gttactcgcg tcgagggtcc gcccgcgcg tagagcgagc 17700  
gcagcaggct gcgaagtccc gacgggagga agcacacggc gtaccgctg agatcggtgc 17760  
gccgcatcgg caggcccgc tgaccgaacc gtttccgcgc cgctgaaac tcgcccggt 17820  
cgactcagag gtggccgagc gcgagcagga gacgatggcg atgatcgctg gcggcgcgat 17880  
gcgcccagctg atcgctcccc cgcgccagga accgtccag cgtcctgacc atgccttcga 17940  
gcatcacgat ggacgacttc gacatgttg tgctgtcag cctgtaatgc gtcaggggtt 18000  
ccgggatgaa accgaccggg cccttcgcgc agaggcgag ccagagggtc cagtctcga 18060  
cgccgtggat gccggtcgcg aacatctcct gccggagcac ggcggtttg accatgaccg 18120  
acgaagcggt cacggcggtt cgcaggacga ggtcggggag acagcgccc tcgagcgcg 18180  
gaagatcgag cgtgcgcagg acctcgccgg cctcgctgat gacggcgcg cccgtatgaa 18240  
tgaagctcca gtcggcggtg tgctggaaga gatcgacctg ccggcgaatc ttggccggat 18300  
cccacgagtc gtcggcgctg agcagcgcca cgaattcccc tcgcgcccg gcgatggcg 18360  
gattgcgggc gccggcaagg ccccggttg cctgagagat caccgtgatg gcgcccga 18420  
atctcgccag gatgtctgcc gtgcccgcg tcgatccgtc gtcgacgacg atcacttcga 18480  
cgggctgata ggtctgggag agcgcgact gcacggccct ggcatgtag cgctcgccat 18540  
tgaacaccgg gacgatcacc gacacgaggc cggggatatc ggtctgggcg gccggcgagg 18600

gctctcgagc ggcttcagcg gccatggccc tctgcaactgg cgtcgcgagc gaacagcatg 18660  
acgtttctcg aaagccacgg agacaacggc gcgctcttca tccgctcccg cagcgtctgc 18720  
gatgccgggc cgaggtagct gaaccccgct gccgcgaaga tcttctgcca gtactcgagc 18780  
ggctgttcgt tgatgtggcc cattccgccc tggcccggct gcgcggcgct gaagagcacc 18840  
agcggcgccg tgttcgtgag cacggcgacg agcttctcgc caagcggcgg cggaaggtgt 18900  
tcggcgacct cgaggctcat ggagaggtcc gttgcgtccg cggcgtgga caccgccgtg 18960  
agatcgaatg agttcagcgc gacgccgagc ttttccgagg cgtaccgtcg cgcgtactcg 19020  
gatcgtcgt agccggccgc cttcacgcct cgtgcgagga acgctgcac gaagtgaccg 19080  
gtgccgcagc cgaagtcgtt catcgatcgg atgccgggtc gcaacgccag gaccccatca 19140  
acgagcacgg gcgcgcctc ggccggcagc tgccgaaca gatccatgac ggccgcgttg 19200  
tacgtccatc gtcgaagcc cagccgttcg ccgatccggg acatcttcag tccggcctgg 19260  
caacgccccg accagtagag ccgcttgagc tgggtcacga tcatgggttc gtcccgggcg 19320  
cgacgcgaac gcgcggcggc gcatgcagcg acgcccagaa ctccacatcg gccacgccgg 19380  
cccacgatcg atgttccttg accgccagcg tcgcaagctg ccgcgacggc acgaggaacg 19440  
tctgggtcgg gttgtgaaac acgtgcaccg acaggtagta gtggccgcga acgagatgcg 19500  
cgcggaatg gaggtcgagc gagaactcgc ccatgaacgc ggacgcgtcc ccgagatccc 19560  
gccggtggat gttgccgtcg tagacgacga gctgatcggg ggagcgttgc accatgaagc 19620  
cgaacgtgac gtcgtccagc ggctcggtag ttctcgcggg gacctggagg ttcagctgct 19680  
tccccggcgt cacttcgctg tgcgtgcccg acgactcgtc gccgacgaag agcgagccgg 19740  
acacgatctc gacggggctc ccgacaactc gctgttccga tttcctggac gcgcggacgt 19800  
acgcgtcgag gacttcgatg gccgtgccgg cggccttcac cgagccggcc agatacaccg 19860  
catgctcgca gagggtcgag accgcctgca gggtgtgcga gacgaggacg atcgcgatgc 19920  
cctggcgctt gaactcctgc atgcgatcga agcacttctg ctggaacgtc atgtcgccga 19980  
cgctcaacac ctctcgatg atcagcacgt cgggccgcag gtgcgcggcg atcgagaagc 20040  
cgagccgcgc gttcatgccc gacgagtagc gcttgacggg cgtgtccacg aagtcggtga 20100  
cgccggcaaa ctcgacgatc tcgtcgaacc gctcggcgat ctcgaccgt ttcacccca 20160  
tgatcgacac ctggaggaac acgttctcgc gcccggtgag gtcgggatgg aaccccgcg 20220  
cgacttcgat gagcgcgccg gcgcgcccc ggacgatcga gcggccggtc gtcggcttca 20280  
ggatcttcgt cagcaccttg aggggtggcg acttgccgc cccgtttgcg ccgatgatgc 20340  
cgagcgcctc gccgcgggtg acttcgaacg agacgtttct gaccgcccag aactcctgtt 20400  
cgtccagcgc gtcggcggga cgcggttgaa acagcccgcg cacgagcgcc ggcacgaggt 20460  
cccgcaggct gtcatgacgc tcgccccgc ggaacttctt ggagacgttg tcgaagacga 20520  
cggggggccgt cacgtcagac gttctccgcg aattcgaact cggagcgatg gaagaacagc 20580  
cagccgaaga cgagcgtcac gaccgagacg atcgccgtga cggccagcat caccgacggc 20640  
ggctggttca ggagcagcac ggccgcggaag ccgtcgatga tcggcgatcat cgggttgagc 20700  
tgcagggcc actggagctt gccgccacc atttcgaccg ggtagaccgt cgacgacgcg 20760  
aacatccaga cggcgatgac gacctcgaag aggtatttca cgtcgcggtg gaagaggttc 20820  
gccatcgaca ggatcagggc gatggccgtg gtgaatatca cgagcacggc gagcacggcc 20880  
ggcagccaga ggatgttcgg tccgaccggc acgtggtagt agatcatgag cgcgacgagc 20940  
accacggagc cgatcgcgaa gtccacggcg ctacgatca cggcgagaa cgggaagatc 21000  
tccttgggaa agtacacctt cgagacgaga ttctgtgtcc cggtcaggga ggtcacggag 21060  
aagcggagcg ccgaggagaa gaagttccag accacagggc cgagaaacgc gaacaccgga 21120  
tacgcgaccg gcgtgtcgat ggccgcgacg cgcgatgaaga tcacggagaa gacagccgtg 21180  
ttgatgagcg gcatgaacac ggccaccgcg aagcccatca ccgactgctt gtagcgcagc 21240  
agcaggtcgc gccgcgtcat ctgcatcagc agctcgcgat actcgaactg ctccggcagc 21300  
atctcacggc aatcgccag gaaggcgcgc atcgtcagcg gtgtcagtcg cggcgtcgtg 21360  
ccgggcggat catgagggca ggtagcgtc gagctgcccg aagagcaact gcacgaactc 21420  
ggtcgcgcgc cgctcggcgg cgacgtcgcc gttttcgtc gacggcacgg ggctcgtcac 21480

gcggacgagc gcggcgctccg atcgggttcag ccgcatcgcg tcgtacacca tgtagatctt 21540  
gctccagtac tcgctggcga tcacgcggcc gtggctctgg taccagtaca gcacgatttg 21600  
ccgctcgatc cccttctgga tgacgtatcg gttgatctcg atgggctcgg tcgggcccgg 21660  
caccttgacc agcgcgcggc cgggtggatac gggctcccag ccggcgcccg gcaggcagtt 21720  
catcggcgag tggatcgtgt cgccttcgcg ctggctctgg tagtagccga tgtagagcga 21780  
cacgtagggc tcgcgcggcg ccaggtagat ccggttgatg tactcgtcca cgcgcagcac 21840  
cgccatgacg tccttggtga agggcgcgct gtcgcgtccg gtccaccgct ccatctggaa 21900  
cgggacgttc gcgagcggct ggcgaagcgg gacctgctcg gcgcgcacatc agcgcgcggc 21960  
aaacgcgcgc gtgccgagga aacacacgct gagaatgatc agtcgtctgg tcatgaatct 22020  
ggactcacac cgtggccggc gcggtcaccg ctccggggcg aggcgtgaag cgcgcgagga 22080  
cccagaccac gacgaacaac atgacgaacg cggcaagaaa gacgagccag cccgagaacg 22140  
tgtggaagaa cccctgcgcc gcggcctcgc cgtagtagtg cgcgcgccag cccgtgccgg 22200  
ccacacgcgc gccgttcgcg ataatcgcg tcgggatcgt cgagagcgca atcgccaggc 22260  
gcatgccgac gcgcggatcc gtgaagtagc cgtacacgat gccagcgtg agcagcgaga 22320  
tgagcgaccg gatcccgtg cagcctcgg ccacttcgag cgtcgtgtgc gcgagcacga 22380  
tcacgttgcc ttgcgcgagc accgggatgc tgagcgcgga cagcacgccc tcgcccacgc 22440  
gagacgccag gagctgcagc ggaaatgcga tctgattgaa gatgatcgac gggatcggga 22500  
tcatcagcac caggaacgcg agcggaaacg cgagcaccca gagcttgcg cagccgagca 22560  
ggaacacgat cgcgcgggta agcagggcta gtatcgacac gcgcgtcaga aagagctcgg 22620  
cgccgagcaa ccccgcgacg agcagcgcg gcccggcgag cacgatgacg agaccgagcg 22680  
cgctcggcgt gtccggcagc gccgagagac gcgcccggcg ctcccaggca aagtagggcg 22740  
cgagcggtag gatcagaatc ccgtgggagt agttgtcgtc gttgatccag tcctgcgcca 22800  
accgcacgat cacctgccca taaaggagca ggaaactcac ggcggcgata ccgagcggaa 22860  
ccagtctggc ctgaagcaca ctcatagcga tcgggacgaa tatagcaaag ggctctcccg 22920  
ggcagaactt gggcgccgaa acacgtaacc atcacgctca caggaagtta gcggctaaat 22980  
gacgggattg gccggctgga ggggcctcgt gaatctaccc aatggatgac aaatatccct 23040  
atcaactcct ccttttgact gcgtagcacg gaagatcatc ggctcttca gatgcaggac 23100  
aaatgtcatc tgtcaggaag agtaatgtct ctggccacag cgcgggtcaa gtgaaatccc 23160  
ccgacagaat cgcgtaccac cgcgactggc atccggcgag aaaagtcgcc tcgtcggccc 23220  
gaaaactgca ttacaggaac gccaggctcg cagccggttc gttttcgtc agtatttttg 23280  
ggcattgcag gcgctttcgt ctgccggggg gcggaaccgt gcgaccgggc cgatcacgcg 23340  
tgagcaggg cagggacag catggatgg tccgtgagac agccgcgttc gggagacggc 23400  
cgggctcggc gggagctcgt ccacctggc gtggaccgcg cggcgcgcca atttctgac 23460  
gcgcggggcg taagttttga cggcgtctcg ctctcgtatc ggggcctgac tcgcccgcgc 23520  
aacgcgtcgc cgcaccgcct gcagcgtcat ggcgtcgate ctgacgtggt ggtcggcatc 23580  
tcctcgaacc ggtccatcga gctcatcgtc gggatcctcg ccgtgctcaa ggccggcggc 23640  
gcctacgcgc cgctcgatcc tgcgtatccg aaggaccgc tcgagtacat ggtcagggac 23700  
gccggggcgc ggatcgtgct cacctcgcga cgcgtgagcg gtctgatccg cgtcgatggg 23760  
gtcgagctca tctgctcga tgcggacgag ccoggggcgt tggcggaaga gggccatgcg 23820  
ccggacgccc gcgtcgatct cgaccacctc gcgtacgtca tctacacctc gggatcgacc 23880  
gggcgcccga agggcgtcgc gatgccgcat cgtccgttg cgaatctcat cggctggcag 23940  
gtcgacgcat cggccgtcgg cgcggcgcg cgcacgctcc agttcacttc gccgagcttc 24000  
gacgtctcgt tccaggaaat ctttgcgacg ctccgcgacc gcggcgagct cgtgctcgtg 24060  
tccgaggaca cgcgcgggga tccgcgcgcg ctgctgcggg tgctgcgcga tcgatccgtc 24120  
gcccggctgt atctcccgtt cgtcgcgttg cagcagcttg ccgtcaccgg caccgacgat 24180  
ccggatctgc cggcgttgcg tgaggtcatc acggccggcg agcagctccg ctcgacgccc 24240  
gcgctcgtcc agtggttcgc gcggcatccg cagtgcacgc tccacaacca ctacggcccg 24300  
gccgaagcgc atgtcgtcac tgcgcacacg ctgactagcc ccccgcgga atggccggcc 24360

ctgccgccga ttggcgtgcc gctgcccgcc gtcgagattc tgatcgtcga cgagcatcgc 24420  
cagcccgtcg cgcacggcaa cgtgggagag atcctcattg gggcggtgtg cctcgctcgc 24480  
ggctacttga atcagccgca ggcacgggcc gaacgggttc tctcgcatcc gcttcgcccc 24540  
gatgctcgcg tgtatcgac cggcgatctc ggccgtgtgc gcgacgatgg cgcgatcgag 24600  
ttcgtcgggc ggacggacga ccaggtgaag attcgagget tccgtgtgga gccgggagaa 24660  
atcgagaccg tcctccttca gcatcccggt gtccgggaga ccgccgtcgt cgcgcacgac 24720  
gacgcgtcgg gcagcaagag gctggtcgcg tatgtgattc cgtcagccga tcaggcggcc 24780  
accggagccg aggcttcggc acaggtctcc ggatggcaga ccgtgtggga ggacacctac 24840  
cgcgcggcgg cggactcgac ggatccggcg cttgcccgca ccggctggcg gagcagctac 24900  
acgggggatgc cctacaccgc ggcgagatg cgggagtggg ccgatgagac cgtcgcgcgc 24960  
atccggtcgc atgcgccacg cgaggtgctc gagatcggct gcggcacggg catggtgctc 25020  
ttccgtctcg cgcctcactg cgtgtctat agcgcgacgg atgtctctcg cgcgcgctc 25080  
gaccacgtcc gggcccacgc gacggcctgc ggccgagcc acgtcgagct cctctcgcgc 25140  
gaggcgccgg actttaccgg cgttcggcg gccgcgttcg atgccgttgt gatgaactcg 25200  
gtcgcgcagc acttgccctgg cgtggactat ctgacgcgcg tcctcgccgg cgcgcgccgc 25260  
accgtgcgtc cgggcggatt catcttcgtc ggcgacgtgc cgaacctgcg cctgctcgag 25320  
atgtttcacg cgtcggtcga gctcacgaaa gcgccggcg atcttccgat tgccctcgtg 25380  
cgcgagcgcg tgcagcgtca gatggcgctc gagcgcgagc tcgtgatcga tccgatgttc 25440  
ttcgacgccg tcatggaccg catccccggc atctcgggag ccaggtcca gatccggcag 25500  
ggtgccttcg acacggagat gaaccgggtc aggtacgacg tcgtgctcac ggtgggcgga 25560  
cggccggccg ggcccgctcc gtccgtcgcg ctcgactggc gtacggagcc ggtcgccgag 25620  
cgtcagctcg ggcaactgct cagcgccgat gcgccggacg cgatcgagat caggaaactg 25680  
accaacgcgc ggctgcgtca gccggccgcg tgcctcgatc tcgtcaagcg tgccgactgg 25740  
accggcacgg ccggcgagac gcttgagagg gccgcgcgg acacgcgcga cgcgatcgac 25800  
ccggaatcgt ggccggcgtct cgcgaacgcc gccggctacg acccgagtt cacctggtcg 25860  
gccgagatcg gcgccatcgg cttcgatcgc cggttcgagc gccggggcac cgcgtcgcctc 25920  
gccgctcccg tgactcgtca cgcggttccc gcccgggagc cgcggggccat tcagcagtac 25980  
gcgacgaatc cgtcccgca tgcggtggcg cggcgccctca cgcggagct tcggcgattc 26040  
ctgcaggagc gcgtgccgga tcacctggtg ccgtcggcct tcgtcgtact cgatgcgatg 26100  
ccgtcacgc caagcgggaa gatcgatcgg cgcgcccttc cggcgcccga gagccgcgt 26160  
cccgatctgg acgtggcggt tgccaagccc gccacggagc tcgaacgcaa gatcgcgag 26220  
gtctggcaga cgacgtcca gatctcgtcg gtgggattgc acgacaactt cttcgatctg 26280  
ggcgggcact cgtcctcgtc cgcgcaggcg ttcgagcgta ttccggcgct cgtgccgggc 26340  
aagcagtggc cgatgatcga gatgttccag taccgcgcgc tctattcgtc cgcggggttt 26400  
ctcagtgage ggccggacgc cggcgccggg cagcttgccg gcgcgcagga tcgcggacgc 26460  
cgccagcgcg agatgctgac gcgtcagggg cctgccctga gcagagtcga aggggtccgg 26520  
taatgcagaa cggcatcgcc ctgcgtcggga tggccggccg ctttccgggc gcgcggtcga 26580  
tcgaggagta ctggcgcaac ctgcgtcggc gcacggagtc gatcacggtc ttcaccaacg 26640  
aggagctggc gacggccggg atcagccaga gcgatctgca gaatcccagc tacgtccgcg 26700  
cgcgcgggct cgtcccggat ccggacctct tcgacgcgcg gttcttcggc atcagccagc 26760  
gcgaagccga gctgatggat ccgcagcacc ggctcttccct cgaggcgtgc tggctcgcgc 26820  
tcgaacacgc ggcgatcgcg ccgcggtcgt tcccgggtct catcggcgtg tggggcgga 26880  
tgagcacggg gatgacgaac agcacgtacc tgctgtcgaa tctccacagc caccgggat 26940  
tgctccagcc ggaggacgtg ctgcggcgcc tgctcggcaa cgagaacgac tacctcacga 27000  
gccgcgtctc gtacaagctg cacctgcgcg gaccgagcat gaacgtgcag accgcgtgct 27060  
cgacgtcgtc cgtcgcgggtg gtgcaggcgt gtcagtcgct gtcacgtgg cagtgtgatg 27120  
cggcgtcgc cggcgcggtg tccgtgtcgt tcccgaaca ggaaggctac gtctacgtcg 27180  
aggcgggcat cgggtcgcca gacgggcact gccgcgcgtt cgatgcgaac gcccaaggca 27240



cggtcttcag caacggcgtc ggcacgtcg tcctgcgaag gctcaggac gcgatcgccg 27300  
acgggcagac gatctacgcg gtgatccgcg gatcggcgct gaacaacgac ggctcggcga 27360  
aggctcagttt cgcagccccg agcgtggacg gccaggcgga ggctcgtggcc atggcgcgaga 27420  
gcgcccgcga cgtggcgccg cagacgatcg actacgtcga ggccgacggc accggcacgc 27480  
cgctcggaga tccgatcgag gtggctgcgc tcacgaaggc ctttcgcgcg ggccggcgcca 27540  
cggagaacgg gttctgcgga ctccgatcg tcaagagcaa tttcggggac ctogactcgg 27600  
cggccggcgt cgcggggcgt atcaaacgg cgctcgcgct ccatcaccag cagattccgg 27660  
cgacgctgca ctacgtcgag cccaacccga agtgtgattt cgccaacagt ccgttcttcg 27720  
tgaacgccgc gctgcgacca tggccgcgcg gtggccatcc gcgacgcgcg gccgtgagct 27780  
cgttcgggat cggcgggacc aacgcgcacg tcgtgctcga ggaagcgccg cccgcgcgg 27840  
agcgaccggg cgcggagccc gccagcgacc tgctcgtgat ctccggcgaag acgcccggcg 27900  
cactcgacgc ggccggctgc aacctcgcga gtcactcga ggccgacccg gacgtcagca 27960  
tcgcccagct ggccgacacg ctgcgcgcg gccgacagga gtttcgcgat cgacgagcgc 28020  
tcgcctgcac cagccgcacc gatgcgatcg ccgtgttcg cgccgaggac gccaaacgtg 28080  
tcgccaccgg cgacccgaa tccggcagcc gtgcccgtggc cttcatgttc tccggcgcg 28140  
gcacgcagta cgcgaacatg ggccgggagt tgctgcgggt cgccgaggtc ttccgcgaag 28200  
aaaccgagcg ctgcctcac cacctgtcac tctccgcga cgcccgccgt gcgtgtttc 28260  
cggatcatgc ggacgtcgag cggccgcgcg attggctcga agctccgtcc gttggtttgc 28320  
ccgcgtctt tgctcgtcag tacgcgtcg cccgtcagtg gatggcgtgg ggtgtgcagc 28380  
ccgacgggct catcggtcac agcatggcg agtacacggc cgctgcctt gcgggcggtga 28440  
tgtcgtcga gaacgcgctc gcgtcgtgt gtctgagagg ccggtgttc gagacgcttc 28500  
cggccgggga gatgctcggc gtcccgttc ccgaggcaga ggtgcggccc gcgtcggcg 28560  
atggactcgc gatcgccgcg gtgaatgcgc caagctcgtg cgtcgtttcc ggccgcccgc 28620  
acgcgctcgc ggcctttgcg cgcgtgatgg ccgaggccgg tgctgaatgc cgcgcctcc 28680  
acatcgcggt cgcggcgcac tcgccgatgg tcgatccgat cctggcgag ttccgagcgtt 28740  
tcctccgcgg catcgcgctc gacgcgcaa ggttccagt cgtgtcgaac ctccgcggt 28800  
cgtggctgac cgcggacgaa gcgaggtcgc cgcagtactg gacgcggcat ctgagacaaa 28860  
ccgtcaggtt cttcgacggg ctccggcacgc tgctcgcga gccgaatcg gtgctcctcg 28920  
aagtcgggccc cggacagacg ctgtccggat tcgcacggca gcatccacag aaaacggcg 28980  
cgcacgtcgt tcaagcgacg ctccggcacg ccaaggaacc ggcccgagc gtctgttcc 29040  
tgcagcaggc ggtgggcagg ctctggacgg caggggtgcc cgtgaactgg gctgcttaca 29100  
acagaggcga gggcaggcgt cgtgtgccgc ttccggccta tccgttcgag aagaagcgg 29160  
acctggtcga atcgccccg gcgcccgtcc cagcgccgt tcaagccctg gcgccaagg 29220  
ccgatgctac tcagcagatg acgacaactg tgccttcgtc tccggcccga tcgctgccc 29280  
atcgtatccg cgagcggctg accgaaatcc tgcacaaact gagcggcctc ccggcaggcg 29340  
agatcgatcc cgcgctgtcg tttctggaga tgggattcga ctctgtgttc ctgacgcagg 29400  
cgagtctgcg gttcaaggcg gattcaagg tccggatcac gttccgccc ctcttcgaag 29460  
aggcaccgac gatcgacgcg ctcacgagct acatcgacgc caagctgcct gcggacgcgt 29520  
ttcctgcgcg ggccggaagc gctccggcgc cttcgtggc cgcaagcgtg ctgtccgcg 29580  
gccctcatgc ggccctggct ggccgtgccg cgcccgccc cgccacgctg cagcacgtga 29640  
ttcaacagca attgcagttg atggccgagc aactccgcgt gctgggccc aatcccggc 29700  
tgctgcccgc gatgcccgcc ccagcgtcgt ccgtcgcggc ggccgacgcg ggactgaagt 29760  
cccgcgtac gacggcgccc gtctcgaagc cgcgtgcggc cttcaagccg atcgacggca 29820  
aagccagtgc ggagctgtcg cccgagcaga tggccgcgct cgccgcctgg atcgatcgat 29880  
acacacggcg cagggccgga tcgaagcgac gcctggccga gtaccggccg gtgcttgccg 29940  
atccacgctc ggtcgcggga ttcaaccggc tctggaagga gatgatctac ccgatcatca 30000  
ccgaccggtc ggaaggatcg aaggtctggg acgtcgacgg caacgagtac gtggatctcg 30060  
tggcggggtt cggctcgtt ctgttcggac accgttcgcc gatcgtggtc gacgccgtgg 30120

aggaacaact ccaccggggg ttcgagctcg gaccgttgcc gccgctcgcc ggcgaggtcg 30180  
cggcgctcgt caaggaattc acggggatgg agcgcgttggt attctccaac accggctccg 30240  
aagccgtgct ggcggccacg cgttttgccc gaacgggtcac cgggcgcgac aagatcgccg 30300  
tcttcgaagg gtcgtatcac ggtctccttg acgaagtgtc cagccggccg ctctgtgtga 30360  
acggcgaatt gcggtcgcc cccgcggcgc ccggcatcgc cggcagcgcg gtgagcgaag 30420  
tcctcctgct cgagtacgcg aaccggaagt cgctcgaggt catccgcgcg cagcagtcgg 30480  
agatcgccgc gatcctcgtc gagcccggtc agagccggcg gctcgacctg cagccgaagg 30540  
agttcctgca cgaggttcgc cgcgtggcgg acgagatcgg cgccgcgctg atcttcgacg 30600  
aagtcatcac cggcttcgcg gtgcattccc gcggcgccca ggcgcacttc ggcgtccgcg 30660  
cggatcttgc cgcctacggc aaggctcgtc gcggcgccat gcccatcggc atcgtcgcg 30720  
gcaatgcgaa gttcatggac gcgttcgacg gcggccactg ggagtacggc gacgggtcgt 30780  
tcccggaggt gggggtcacg tactttgccg gcacgttcat gcggcatccg ctgcgcgtcg 30840  
ccgctgcgaa ggcggtgctg acgtacatga agtcgcaggg tcccggcctg cagacgcggc 30900  
tgccagatcg gatcgagcgg ctgcgcaacg acgcgcgcgc gatcgtggcc aggcacggcg 30960  
cgccgtatca gattacgcag ttcagctcga tgatgagcct gaacttccc cagcaccaga 31020  
agacggcggg gctgctgttc ttctgatgc gcgagcgcgg gattcatatc tgggagggcc 31080  
gtcccttctt cttcacgacg gctcacaccg aggcgcacta cgccgccatc ctccgcgcgc 31140  
tcgacgagtc gctggccgag atgcaggcgg ctggcttctt cggcgcgccc gccactccg 31200  
cattgcaggt gaccgggacg gcggaccagg acgtcctgcc gttcaccgaa ggacagcagg 31260  
agatctggct cggcggccag atgggcgagg cggcatcacg cgccataaac gaagtcgtcg 31320  
cgctcgacct gcgcgggcct ctgaatcgcg cggcgatgca gcgcgcgcgc gacgaggtcg 31380  
tcgcgcgtca cgaatcgtc cgcattgacg tggccagcga accgctcggc ctgcgtcacg 31440  
ttccgggcac gtcgggtgcc gttgaatggc aggacgtgtc gggattgccg gaggacgctc 31500  
agcacgaggc cgtgcagaag gtgctcagg gcggcgacgg cgtggcggtt gacgtggaac 31560  
gcgggcccgt gctgcgcgtg accaccatga cgctcacc cagacatcac gtcttcgtca 31620  
tggcgcccca ccattctcgtg tgtgacggct ggtcgttcgg cgtcatcctg cgcgatctgg 31680  
cgacgttcta ttccggcccag gttcgaggca cgcgtgcgaa cctcgaggcg ccgatgcagg 31740  
tgagccgctt cgcgcgtgag gatcgcgagg ccaagcagag tcctgaagcg gccgagaccg 31800  
aggcgttctg gatccggcaa ttcgacaccg tcccagagcc gctcgagtta cccgccgatc 31860  
ggccgcggcc gccagtgaag tcgtaccagg gcgcgcgtgt gtcggtgccg tttgacgcgg 31920  
cgctcgcctg cggcgccaag aagctcgccg ccgaacaccg cagcaccctg tttacgacgc 31980  
tgctcggcgt cttcgagacg ctctgtctacc gactgaccgg ccagcaggat ttcgtcgtcg 32040  
gcatcccgtt cgcggtgcag cctctgctcg gcgaggatct ggtcgcgcac tgcgtcaact 32100  
tcgtgccgtt tcgcgcgcgc gtgtcaccca ccgcgacgtt ctccgagtag ctgaacacgc 32160  
tgaggacgca gtcactcgac ctcaatgagc accagaactt cacgtacggc agtttgctcg 32220  
ccaagctgcg gttgccgaag caccggggcc gcgatccgct cgtctcgacg agcttcacgc 32280  
tcgagccgc catcgtcgat cccgggttcg acggtctcac ggcgcgttcg ctgacggttc 32340  
cgcgcgtgac atcgaagcgc gacctgcacg tgaacgtcat ggagtggac ggcggtctgc 32400  
tggtcgaggc cgactacagc accgatctgt tcgacgaacc gacggtgcgc cgggtggctgg 32460  
atagctatcg gatcctgctc gaggcgctcg tgctgtcccc cggccgggtc ctctgagcg 32520  
tgccggtgat ctccgagccg gaccggagtc agcttgtcac cggctggaac gacacggccg 32580  
ccgactaccc gcgcgatcgc gtcattccatc aactcgttga agagcgcgcg gtgcagaccc 32640  
cggagggact cgcgctcgtc tgtgggtgcc agcggctgac ctggaccgag ctcaaccgac 32700  
gcgcaaaccg gttggcgcgt gcgctcgtga agaaggcggt cgccccggg agcaggggtg 32760  
cgctgtgcct ggaacgatcg gccgggttca tcgtctcggg gcttgccgtg cacaaggccg 32820  
gcgccgcata tgtaccactc gaccccgctc cgcccggcga tcgcaagtcc ttctcgtcg 32880  
aggattccgg tgccgtgctg gtcgtgaccg attcgcggtc cgaaggccga tggctgacgc 32940  
ccaggacgcc cgtgctccac ctgcagccg acagtccgcg aatcgcgcaa gactcgacg 33000

atgatctgaa ggtcgccctg tcggccgaag acgccgcgta cgtgatctac acgtccggat 33060  
cgacggggaca accgaagggc gtcgtcatcg ggcaccgcca gctcgtgaat tacacgtggg 33120  
cgggtgatcga acggatcggg ttacccgccc gctcgagtta tgcgctcgtg tcgaacgtcg 33180  
cggccgatag cgggtgtgacc gggctctcgg cgtcgcgat gggctgggtg ctgcacgtga 33240  
tcacctcaga ggtcgcgacc gacggggcggg cgtcgggtc gtatttgacc gttcacgaca 33300  
tcgacttccct gaagatcacg ccgtcgtcgc tcgctcgtc gctcggcgat caccgctcgg 33360  
cctcgtcgtc gccgcgccgc tgtctcatgt tgggcggcga gccgtcacgt ccggcgtggg 33420  
tggaagagct gcagcgtctc gcaccaggct gccggatcat gaaccactac ggcccgcgag 33480  
agaccacggt cggcgtgttg acgtactgga ccgggtgacc cagggatctg ttcgtcgaga 33540  
gcgtgccggc cggcaggccg ctcggaacg tccgcgccta cgtgctcgat ccggccggag 33600  
cgcccgcgcc gattggcgtc gtcggcgagc tctgtatcgg cggcgcgtgc gtggcacgcg 33660  
ggtatctcaa tcgtcccga ctcaccgcc agcgttctgt ccccgatccg ttcgccaccg 33720  
agccgggagc gcgcatgtac cggaccggcg accgcgcccg gttccgggcc gacgggaaca 33780  
tcgagttccct cggtcgcgcg gactatcagg taaaagtacg cggattccgc gtcgagctcg 33840  
gcgaaatcga agcggcgctt cgcgctcatg agggcgtcga gcaggccgta gtcgtgctgc 33900  
gaaaagatca gcccggcgac gagcgcctga tcggctatgt cacgaccggc ggcggcgcct 33960  
cgatctcgat ggcggagctt cgcacgtcgc tcaagcagag gttgccggac tacatggtgc 34020  
cggcgggtgat cgtcgcgcgc gacaagctgc cggtcacgtc gattggcaag atcgatcgac 34080  
gogccctgcc ggcgcgcccc gagcgcgcca gtttcgagtc ggagttcgtc acgcccgcga 34140  
ccgagacgga gcgcgggctt gcggagatct ggagcgcggt gctcgggtctc gagaagattg 34200  
gcgcgcacga caacttcttc gacctgggcg ggcactcgtc gctcgcggcc cgcaccgcca 34260  
tgaagggtccg cgacacgttt cgcgtggacg tgcgtcatc cacgttcttc gaaatgcccc 34320  
ccatcgccat gatggcggca gcaatcgacg ccggatccgc ggcccctgcc tcccggacca 34380  
tcacgccgag ggcccgcgtc agggcccgca gggatgagct ggtccgcccc tgaaccccgg 34440  
ccaccgcgag actgagtcct ccgacctgcg gcgcgacggg atcgccgccg aagcgggcga 34500  
cgtgtttctg gcgcggcggt cgtacgcccc ggggcgcttg tggttccttg cgcagctcga 34560  
gcctgaaagc tcgcctaca acattccgat cgcgctcaga ctgaagggcc gcctcgacct 34620  
cgccgcgtc gaacgcgcca tcaacacccat cgtccagcgt cacgagacgc tgcggacgac 34680  
gttctcgtg cagggcgatg agctcaagca ggtcatctct ccgacgctgc gtattccgat 34740  
ctccctcgtg gacatgcgtc tgctcagtc tgagaaacgc gaggcggat ggccgcgcgt 34800  
tgccggcgac gacggcggc gcgtcttcga tctggctgag ggcccgtgc tcaaggctcgt 34860  
actcgtcatc ctcgagccct ccgatttctg cctcctgac accgtgcac acatcgtctt 34920  
cgacgggttg tccggcggga tcttcacccg cgagctggcc gaggcgtacg cggctgagct 34980  
cgacggggcg gaccgaagc tgccggagct gccgatccag tacgoggact ttgccgcgtt 35040  
ccagaaggaa cagaccgaag ggcccggcct gccggtgctg ttgaagcaact ggacgcagga 35100  
gctggcgggg gcgtcgatga cgtcgcacct gccgacggat cgtccgcgtc cgcctcgtcga 35160  
gacctcccgg ggcgcctcgc cctcgttctg gctcgacgga gcgatctggc cgggcgtcgc 35220  
cgcatcagc cggggcgaga acgcgacggt gttcatgacg cttctctccg ccttcggcgc 35280  
cgtgctgcat cgtggttccg gacaggaaga catcctcatc ggcagcccgt cggccggccg 35340  
cgatcgtccg gagcttgagg gcctgatcgg gttcttcac aacacgttcg tcttcgggt 35400  
caacgcgtcg ggagatccc cgttttcgga gttcgtcggc cgggcgcgcc gcgcgtgccg 35460  
tgccgcctac gcgcatcagg acctgccgtt cgatcggtc gtcgagacgc tcaaccgga 35520  
gcggaccggg gatcggtcgc cgatctatca agtcatgttc gcgcaggatc cgcggcgctc 35580  
gaagatgacc ttcgccggga tcgagctcga gcggcttgtc gtccacaacc gcagcaccaa 35640  
gtgcgacctc acgctcgagg ttgccgaaga gccgacggc gtcaccctgt acgtcgagta 35700  
cagcgtcgac ctgttcgatg cggcgacggt gggtcgtcgt ctgcagcaat atcgccgtgt 35760  
gcttcagacc gcgatcgcg acccggaaca gcggctctcc gagctgcgtc tgcttgggga 35820  
cgacgagcgc atgcagctga tcgactggg gaccggtccg gcggccgtct acccgcgcca 35880

cgtcagtcctg gccacggtct tcgaagagcg ggtccgcgcg actccggcgc cagtggccgc 35940  
gacgctggag ggacagcacc tcacgtacgc ggatctcaac cggcaggcga accgcctggc 36000  
ccggcggttg aaggcggttc gtgtggggcc gaacgtgctg gtcggcgttg cgctcgaccg 36060  
gtcgttcgat ctgctggttg cgctcctcgc cgtggcggaag gccggcggcg cgtatctccc 36120  
gctcgatttc gctcaccgc aggaacggct tgcgttcattg ctccgcgaca cgcgcgcgcc 36180  
catcgtgctc acgctgcggc gatttgccgg ggcgctcgag tcttttcccc tcttctgcct 36240  
ggacgacgag ctgccggcca ccgccgagga gccggacgaa gatctgccgc cgcagtcgac 36300  
cggagaccac ctggcctatg tgatgtacac ctccgggctcg actggtcagc ccaagggcgt 36360  
cgcgatcacg catcgcggcg tggttcgtct cgtacgcggc acagactacg tccgctggaa 36420  
cgacgtgcgg gcgacccctc acatcgcgcc gacgtcgttc gacgcgttga cgttcgaggt 36480  
gtggggcgcg ttgctgaacg gcgcgcggct ggtgctggtt cccgagcggc tcgtcagcct 36540  
ggagacgctc gagcgcacgc tccggctcga gcaggtggac tgcctgctgc tcacgaccgc 36600  
gctcttcaac gccgtcatcg acgcgaagcc cggcattctc gtcggcgtga agcaactcct 36660  
gatcggcggc gaagcgctgt cagtcgcgca cgtgcggcgc gccctggcgc acctgccgga 36720  
cacgcggctc gtgaacgcgt acggctccgac cgagtgcacg accatcgctg gcgcctacca 36780  
gattccgcgc acgctcgatc cgcaggcgcg gtccatttcg atcggccggc cgattgccaa 36840  
tacgcaggcg ctcatctctg atcgacacgg cgacctcgtg ccgttcggca tcgccggcga 36900  
gctctgcctc ggcggagacg ggctggcgcg ggaatacctg aatcagccgg cgctcactgc 36960  
ggaacggttc gtgcccacg cggttcagccc cgagccggcc gcccggtgt accgcacggg 37020  
agacctcgcg cggatgcgcc gcgacggcaa catcgagttc cttggccgga tggatcgcca 37080  
gatcaactg cgcggcttcc ggatcgagcc gggcgagatc gagcatgccc ttcgtgcccc 37140  
gggagacgtt ctggacgcgg tgggtggagt ccgggtcgat gccacatgg cccatctcgt 37200  
cgctacgtc gtgcgcgcgg atgggagtca actcaccgga accgacctgc gcgaacgtct 37260  
gaaggcgcgc ctgccggaat actgtgtgcc cgcgaagtac gtgttcctgg ccgaggtccc 37320  
gcgaacggcg gccggcaagg tcgatcgctc ggcgctgcct gcacggcgga tcgaggcgcc 37380  
cgcgcccag ccgcgcgatc cgtcgttcct gaacgaggtc gaagcggagc tcgcgcgaat 37440  
ctggagcgaa gtgcttcggc tgccctcctg gtcgcgacg gacaattttt tcgatcgtgg 37500

<210> 2

<211> 37507

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: DNA sequence  
of clone FS3-135.

<400> 2

gatcccacga tcggcgccag cggattgcgc agctcgtgcg cgagcatcgc gaggaactcg 60  
gtcacgtggc ggccctcggc ctccagctcc tccatgcgct tgccgtgcga gaggtcgcgc 120  
gtcaccttgg cgaagcccat gtgccttcca ctgcgctcgc gaagcgccgt cacgacgacg 180  
tcggcccaga agcgcgtgcc gtccttgcgc acgcgccagc ccgtgtcctc ggcgcggcgg 240  
tgacgacgag catcgtgcag ctccggcctcc gggcgggcgg ccgcggcatc ctccggcgga 300  
tagaagagcg aaaagtgcgc tccgatgatc tcttcggcgc tgtagccggt gatgcgctcg 360  
gcgccttgt tccagctggc cacgtggccc gcggggtcca gcatgtagat cgcgtagtcc 420  
tgacgcccct cgatcaggag gcgcgagcgc tctcgccttt gccgcagctg ctccctcgtg 480  
tcgcggcgct cggtcaggtc gcgcgtgacc ttgccgaagc cgaccagcgt gccgtcctcg 540

tcgcgcagcg ccgtgatgat gacgttggcc cagaagcgcc ggccgtcctt gcgcacgcgc 600  
cagccctcgt cctcgaagcg gccgtcctcg cgtgcgcgca gcagtcctt ggccggccag 660  
tcgcgcgcga tgcctcctc ggtgtagaac gcggagaaat ggccggccgat gatctcttcg 720  
gcgcgatagc ccttcaactg ttcggcgccc gcgttccacg aacgcacgta gccctcggga 780  
tcgagcacga agatcgcgta gtccgtgacg gacccgacca gcagccggaa gatggcgctc 840  
tcgccggcca gcccgctcgg ggccgccgaag gcgaggttgc gcaccgaggg ccgcggaccg 900  
gatgggcttg gcatggctat tagggtagaa cgggtgcacg aagttgcaat cggacaatca 960  
gagataattg cttgatgaag atgccatttt tggaattctt acgcagggag ctggggacgc 1020  
tcgaggccca gggcctgtac aagtccgagc gcgtgctcga ctctccgag ggccgccagc 1080  
tgaggtggc cgcacacgac gtgtcaact tctgcgcaa taactacctc ggccctggcca 1140  
accaccggc cctggctcgc gcccggaagg agacgtcga gcgcgacgc tacggcatgg 1200  
cgtcggcttcg cttcatctgc gggacgcacg ccgtgcaccg cgagctggag acgcgcctcg 1260  
cgcgcttctt cggcaccgag gacgcgatcc tgtacggctc ctgcttcgac gccaacggcg 1320  
ggctcttcga gacgtgctg ggccgaggag acgcggtgat ctcgacgcgc ctgaaccacg 1380  
cctcgatcat cgacggcggt cgcctctcga aggcgaagcg ctcccgctac gacaacaacg 1440  
acatggcgag cctcgaggcc cagctgaagg cggcggaagc ggccggcgcg cgctacaaga 1500  
tgatcgccac cgacggcggt ttctcgatgg acggcatcgt cgcgaacctg ggccgccatct 1560  
gcgagctggc cggcgctac cagcgatgg tgatggtgga cgactcgac gccgtgggct 1620  
tcattggcg gcacggcgcg ggaacgccc agcattgcg tgcgaaggg aaggtggaca 1680  
tcctcaccg cagctcggc aaggccctg gcggcgcttc cggcggtac acggcgggca 1740  
agcgcgaggt ggtggcctg cttcgcaacc gctcgcgcc ctatctcttc tccaacacgc 1800  
tgatgccgc catcgcgggc gcgtcgtca aggtgctcga tctcctcgag ggccggcgcg 1860  
agctgcgcgc gaagctcgc cgaacgccc gccacttccg ccggcgagat acgcgcctcg 1920  
gcttcacgct ggccggcgcc gaccatccga tcatcccggt gatgctggg gagggcccg 1980  
tcgcgaagga gatggcgga cggctgctga aggaaggcat ctacgtggtc ggcttctcgt 2040  
ttcccggtg gccccgagg caggcgcgca tccgcacgca gatgtcgcc gccatgaac 2100  
cgaagcacgt cgatcgcgcc atcgccgct ttgccaaagt cggcgcgat ttgggagtca 2160  
ttgcttgaga accctttcga agacgaagc cgagcccgcc atctggatgg tcgagtcgcc 2220  
caagcccgtg gtcggccaca acgagtgct gatccgcgtg aagaagaccg ccatctgcg 2280  
caccgacatg cacatcttca actgggacga ctggtcgag aagaccatcc cggtgccgat 2340  
gacggtcggc cagagtagc taggcatggt ggaggccatg ggccaggagg tgcgcggcct 2400  
gcaggtcggc cagcgctct ccggcgagg ccacatcgtc tgcggccatt gccgcaactg 2460  
ccgcgcggg ccgcgccacc tgtgccgaa cagcgaggc gtggcgctga accgccccg 2520  
cgcgcttcgc gactacctg tgattccgc ggagaacgc ttcccgattc ccgacgacat 2580  
ccccgacgag atcgctcga tctcgatcc gttcggaac gcggcgca cgcgctctc 2640  
gttcgacctg gtggcgagg acgtgctgat caccggcgcc ggcccgatcg gcatcatggc 2700  
cgcgcccatc gcgcgccag tggcgcgcg ccacgtcgt atcacggaca tgaacgacta 2760  
ccgcctggcg ctcgccacga agatgggtgc cagccgcgc gtgaacgat cgaaggaaaa 2820  
cctgaaggac gtgatgcgc agctcgcat ggtcgaggc ttcgacgtc gcatggagat 2880  
gtcggcgctg cctcggcct tccgccagat gctcgacacc atgaaccac gcggcaagat 2940  
cgccatgctc ggcattccgc ccagcgagg ccgcatcgac tggaccagg tgatcttcaa 3000  
gggcctggtc atcaaggcg tctacggcg cgagatgtc gagacctgt acaagatgat 3060  
cgccatgctg cagagtggc tcgacctc gcccatggt acgcaccgt tcgacgtgcg 3120  
cgagtacctg aagggtctc agacgatgg ctccgggaaa tccgggaaag tcgtcctgtc 3180  
gtgggattag cgagggccac ggcgcggc ccgggtcccg cgccggcgat cgtgcaggcg 3240  
ggttggcagc acgccccgg atttccgcg tggctcacgt gcgacttcgc cgtcgacgca 3300  
ccgatcgca tcgacgtcga cgtgagcgt tcgctgctc atgccaggg cacctcgtc 3360  
tgacgcagc cgatggcg cagcgggcg gacccccgt ggctaccgac cggcacgtat 3420

cgcgccctccc tgcgcctgga gccgttccgc ttcccgccgc cgccgcctc ggtcgagttc 3480  
gcgctctcga tgcgggtgca aggtgagcgc cgcgtogtgg ccaccgcagc gggcggtggtg 3540  
cccgcgggcg caccacacgc cgtggcgcg cgcgcctggc acctcgaggc cctggatggc 3600  
acgccggcgc tcgagacgct ggcctggtcc gatccacgc acagctggtt ctcccagcat 3660  
ttcgaccatg cgtcgcgcac gtgcgtcgag tacctgggcg ccggtatggc cgggtggcgc 3720  
gggcgcgtgc tggatgcggg atgcggcgac ggcacacgc cgtcgcgcac cgcgtgcgc 3780  
tatgcgccc agcaggtggt gggcatcgat ccggggcgct gctaccgggt gcttcccgc 3840  
atcctcgagc gccacgcct cggccacctc gcgctgccgc acaacctgca gttcctccc 3900  
gcggtatgca acgcgctccc gttcgccgac agcagcttcg acgtcatcgt ctcggtggtc 3960  
gcggtcgagc atttcgtggg cggctacctg ccctcgctcg ccgaggcgcg gcgggtgctc 4020  
aagcccgcg gcctgctggt catccatccc gagctctact acaccgcgc caacggccat 4080  
cacctcgcg agtacagccg cgagccctt ttccacctgg tgaagtcgac cgacgaggtg 4140  
cgcgacatcg tgttcggcgc ggacgtgagc ctccacgacc gcggcgccat cgcgcgcgc 4200  
cgcgcggagc actggcgctg gtacaacgag ctcaatcgca tcacggccgc gcagttcgag 4260  
gacgagttgc gcgcgctcga cttcgagccg tggcgccctg cactgcgcgc cgagccgctg 4320  
gtggagtaca agccggagct gctgcgctac cgcttcacgc agctggcgct ctccgagctc 4380  
tacgtggcct gcatcaaccg caagcccggt tcagatggtt cctacaggac gcatgggcag 4440  
gcggcattcc ggtaaggtcc gccgggacgc gaattgaagg cgcaagcgcg ttgttatccc 4500  
cgtcactccc ggactgcgc gatcaccctc acaggagaac ccccgatgaa gaccagacc 4560  
caggccgtgc tcgccgcgc cctgctgtcc cttgccggca ccacgctggc ccgcgcgaa 4620  
gccgtcacgc tgacgggcgc gagcgaagtc ccgccgtca cctcgagcgc gaagggcgc 4680  
ggcacgggtg tcgtgaaggc cgattgcacg gtcaccgcga agatcacctg caccggcatg 4740  
accgccaccg cggccacat ccacgaaggc aaggccggcg ccaacgcggg cgtggcagtc 4800  
ccgttcgtga agaccgcga cacacggtt gaggcagccc cgggcgcca gatgaccgag 4860  
gcgcagtgcg ccgcgtacaa ggccggcggc acctacgtga acgtgcacag cgaggccac 4920  
aagggcgcg aggtccgcgc ccagctggcc ggcaagtagg gcagccgctg caaaaagaaa 4980  
aagcccgcg aagcgggctt ttctgttgaa ggctttggtc gggcgaaaag gattogaacc 5040  
ttcgacccc tgcacccat gcaggtgcgc taccaggtg cgctacgcc cgaccgaact 5100  
tgaaattata cctgcatcg ggtcgcagcc gccatgaatc gaaacgcccc gggcttcggg 5160  
tagcatccgc ccatggcgac gaccctcgt acgggtggtg caggctacat cggcaccac 5220  
atcctctgcg cactggccca ggccggcgga cgcagcatct gcatcgaaa ctactccaac 5280  
agctcgccgc gttccatcga acgcgtgctg cagatcgac ccggttgct cgaggcctc 5340  
gacgtggaca tccgcgacgc cgacggcatc cgcaaggtga tcgcagggcg cgacgtggac 5400  
agcgtgatcc acctcgcggg actgaaggcg gtcggcgaaat cggtcgagca gcccgagcgc 5460  
taccacgaca acaacgtccg cggcaccgag agcctgctcg cggcgctcgc cgactcgccg 5520  
gtgcgcaagt tcgtcttcag ctctcggcc accgtgtacg gcctcgccga gaagatgccg 5580  
atcgacgagg acgcgcccac ctgcgcgcag agcccgtagc gccagaaca gctcgacatc 5640  
gagcacatgc tggtcgccc cggcaagcgc gaccctcct ggcgcgtagt gaacctgcgc 5700  
tacttcaacc cggtcggcg acacgagagc gcgctgatcg gcgaggacc ggccgggggt 5760  
cccaacaacc tcatgcccta cgtgtgccag gtcgctccg ggcgactgaa ggagctgagc 5820  
gtctacggca gcgactacc caccgggac ggcaccggca tgccgattt catccacgtc 5880  
tgcgacctgg ccgaaggcca cgtggccgcg ctggaagcac tcgagcgcg gggcccggg 5940  
accgtcctca cgggtgaatct cgggacgggg ctcggtata ccgtgctgca attgctcgaa 6000  
accttcgagc gcgtgaacaa actaagcatt gccggcgca tcgtggggcg gcggccgggc 6060  
gacgtcgca tctgctacgc gaacgcgggc cgcgcgaagg ccgtgctggg ttggaccgcg 6120  
aagcgcgaa tcgaggagat gtgcgcgat gcctggcggt ggcaggaaag gaatgcaaag 6180  
ggctacgggt gacggattgc aaatcatggc cttggcgacc cttgctgcaa tgcggcatga 6240  
agggggcggt tcttgaaga tccaaggctt gggatgcgtc caggacgtca aaaacagcac 6300

ccggcggttg aaggaagaag gcaaccctag tgttgtcatg gctccagacg ggattgcgcg 6360  
ccgagtaaga atcgtcagtc gagaccaact cgcctccggg tgcaacgaat cgtgagccga 6420  
agcaacacct cgcacgcgt cgcgcgtcgc ggcattgcgt tccgccttcc cggcgctacg 6480  
cgcaaatccc tctggccctc gctcctcgag gggcgcaacc tcgtcaccac cgtcgatccc 6540  
gcgcgctggg cgaaggagcc gtactaccac ccgcgcaagt ccgagcccgg caccagctac 6600  
tcgttcgcgc ccggcaccat cggcgacatc agcggcttcg acgcgtcgtt cttcggcatc 6660  
tccccgcgcg aggcgcgcga gatggatccg cagcagcgcc tgctgctcga gctctcctgg 6720  
gaagccctcg aggcgcgcgc ggtgaagccc tcgacctgc gcggcagcga ttgcggcggtg 6780  
ttcgtgggccc tgtccaccgt cgactacagc tacggcctcg cctacgacct ggcgtccatg 6840  
gacgcgtcgt cggccaccgg caacaccgcg agcatcgcgc ccaaccgcct ctcgtatttc 6900  
tacgacctcc acggccccag catggtcgtg gatacggcgt gctcgtcctc gctggtggcg 6960  
ttccaccagg cctgccagtc gatcgccatc ggcgagacgc gccaggcgct ggtgggcggc 7020  
atcagcctgc acctgcatcc gtacgggttc atcgcgttct ccaaggcgct gatgctctcg 7080  
aagcgcggcg cctgcgcgtt gttcgacgc agcggcgatg gctacgtgcg ctcggaaggc 7140  
gcgggcgtgc tgggtctgaa ggacctgcgc cgcgcgtgg ccgacggcaa cccgacctc 7200  
gccgtcgtcg ccgcgagcgc cgtgaatacc gcgggcgcct cctccagcct caccgtgcc 7260  
agcgtcgatg ccaggctgc attgctgtcg gatgtgtacg agcgcgcgg catcgaccgg 7320  
tccgagatcg actacctcga ggcccacggc accggcaccg tgggtggcga tccgatcgag 7380  
acccgcgcgc tcggcatcgc gctcgggcgc ctgcgcgcgc gccaccagcc gctgcccagc 7440  
ggttcggtga agagcaacgt cggccacctc gaggcgcct cgggcagggc gggcctggtg 7500  
aaggcgtgc attgcctcaa gcaccgcgc gtgcccga ccatccacct cgattcgccc 7560  
aaccccaaca tccacttcga cgagtgaac ctgaagccgg tctcagagcc gctcacgctc 7620  
catccgtaca agaagctcgt cgtcggcgtg aactcgttcg gcttcggcgg ccgcaacgcg 7680  
cacgtgatcc tcgagagcac gccacttcg cccaaggcgc tcgaggcccc ggccggcgag 7740  
gtggcgctca tgctgagcgc ccacgacgac gcggcgctgc gcgacgtggc ggcctcgtac 7800  
gcgaacttcc tgaggcgag cgagcaaccc gcgctgcacg acatcgcgta caccaccgtg 7860  
cactgccgcg actggcacca gctccgcgcc atggtggtcg gcaccgaccg cgccagcatc 7920  
gccggcgccc tcgcgcaatt cgccaagggc cacaacacgc ccggcgctcgt gagcgccgt 7980  
gcgctgccc aacgcgtccg ccccgcatc gtgtattcgg gcaacggcgc gcaatggccg 8040  
gcaatggggc gcgcctcgt cgcggcgag ccggtgttcc gtgcccgcgt gcaatcggtc 8100  
gacgcgtgt tccgcccgt cgcggcgtac tccatcctcg agatgctgga gcaggaggtc 8160  
cccgcgacc agctggaccg caccgagatc gccagccgc tgctcttcgc cgtgcaggtg 8220  
ggcttgaccg agctgctgc cactggggc atccagccca cgcgcgtcac cggccacagc 8280  
gtgggcgagg tcgcgcgcgc ttgggcgagc ggcgcgctga cgtcagcga tgcggcgcg 8340  
gtgatctacc accgcagccg ctaccagggc gagaccaagg gcgcgggcgc gatgaccgcc 8400  
gtcggcctcg ccgaggacgc cgcgcgcgc cagatcgagg cgtcgggtt gaagggcgag 8460  
gtgaacgtcg cgtgcatcaa cagcacgcgc aacgtgacgc tcgcccggcac gcgcgcggc 8520  
atcagggcgc tcgaggccga gctcacgcag cgcaaggtgt tccaccgccc cctcgacctc 8580  
gactacgcgt tccacagccc ggcgatggac ccggtgcgcg acgggctcgt gagcgcgctg 8640  
cgcgccctca cgcgcgcgc caccgggtc gcgttccatt cggcggtcac cggctcgccc 8700  
gccgcgggca accagctcga cgccacctac tgggtggagca acatccgcga gccggtgcgc 8760  
ttccaggccg cgatccgcgc catcgccgag tcgggcgtca atgtcttcat cgagatcggc 8820  
ccgcacccga tctcaagaa ctacatcaac gacggcctgc gcgcgcctc gatcgagggc 8880  
cgcgcgctgg tcacgtgca gcgctcggcg agcgtcgcgc cggcgatccg cgccgcgcgc 8940  
cagggaagtgc tgatcacggc ctgcccgggtg gagaccgcga agatgtccc cgcgagggc 9000  
cacttcgtcg agctgcgcgc ctaccgtgg cagcgcgagc gccactggcg cgcgcccacc 9060  
tcgaggcct acgacctcat ccagcacggc aagcagcacc cgctgctggg ctatcgctg 9120  
cacgagaacg acttccagtg ggaaaaccac atcgacaccg cgctctatcc ggcgtacgcg 9180

gaccatgtcg tcggcggcgc cgtggtgttc cccgcggccg gcttcgtcga gatggegetc 9240  
gccgcctcgg cgatcgcgct gggcggcgac gcgcacgaga tcgagacgct cgagatccgc 9300  
agcgcgctgc tgctcgagga ctccacgtcg aagaccgtgc gcttcgcgct ggaacccgat 9360  
ggccgcttca caatccgcag ccgcgctcgc ctgagcgagg atccgtggca actccacgtg 9420  
gtcggcaagc tggtcggcac gccaccgag ctccgcgcg cgccgcgcca gtcggcgccg 9480  
tcgcgcaagc ccgatgtgat cgcggccccg cactacgaag gcgcgcgaa ggccggcctc 9540  
gcctacggcc cggccttcca gtcggtctcc aagggtgtggc tcgacgcgc ggaccgggc 9600  
agcgcctacg cgcgcctgat cctccccaag ccgattcgcg gcgagctggg cgtgatgcac 9660  
ctgcaccccg cgtcgctgga cggctgtctc cagctgtctg tcgacctgct gcgcgcgag 9720  
gcctcccgcc acgcgcagat cgccttcgtg ccgattcctg tcgggcgcac gcgcctctac 9780  
ggcctgccc cctcgtcac cgcctgccc gtcgcctga ccgcgcgag ccgcgctcg 9840  
ctggtcgcgg acttcgagct ctacggcccc aacggcgacg tggtgccgc gctcacggc 9900  
gtgcgcttcc gcggcgctcca gctgaagcag ccgcagggcg cgcgcctgc ctacctgtcg 9960  
taccgcgcca tcccgcgct gaattggcggc gacacgcgg gcgccaagcc gctgcccctc 10020  
gcgttgctgg ccgcggcggt cggcgagcgc tggcattcgc tctcgccgg ctcgcgcaag 10080  
cgctactacg atgaggtcga gccgctctcc gacgtgtgt gcagctcct cgccgagcac 10140  
gcgctgcgaa agcggatcgg cgaggcccc ctcacgcacc cggccgcgct cgccgcgtcg 10200  
acgcggccg agcattggcc gtcctcaag cgcgcgatcc agatgctggt cgaggaccag 10260  
ttgctcgagc ccgccgaagg cggctggcg tggagcccgg tcaccgaatt gcccgacgcg 10320  
caggagagct ggatcacgct gctgcgcgac taccggacg aggcggcca gctgatcgcg 10380  
ctcgggcgcg ccggcgctca cctcgcgaa gtgttcgccc gcaaggaagc gtccttgccg 10440  
agcctgttgc cggccacgca gggaacgtg ttcgcgcgcc tgtgctcggg cgcacccgcg 10500  
ttcgccaacg ccgggcgcgc catcgccgac acgctggcg tcgccctcga gcgcctgcc 10560  
gcgcaccggc cgtgcgcgt gctcgaggtg acgcgtgcc gctccgagat cacgtgcgc 10620  
gtcctcgagg ccgtggaact cgaccgtgc gactaccag tcggctgcac cacggacgaa 10680  
gactggggc agtacgaagg cgtcctcgat gccattcgc acgtcgagac ttgcgtcgtg 10740  
gacctgaaga tcccgggcat gggcctgaag gcgccgaacc agggcccgtt cgatgtcgtg 10800  
atcgtttccg atggcctgct cggtgcgcc gatccggat cggccgcggc gcacctcgcc 10860  
ggcgcgctcg cggaagacgg gctgctggtc atggtcgcgc agcatccgtc gcgctggacc 10920  
gacctcctct tcggcgcgga cccggcggtg tggacgtcg ggcctcggg ctgcgagcg 10980  
tcgcgcctgc gttcgccgga cgagtggcg atcgcgctg cccacctcg cttcggggcc 11040  
accgtcgcgg tgcccagact gccggccctc gacgcgggt cgtacgtcct catggcgcg 11100  
cgcgacggcc cggtgccgca ggctgcccag cgcagcgttc ccgccaagtc gggctggatc 11160  
ttcgtgcagg aagcgggcgg ttattcggcg gcgttcggcg cctgcgtgca ggcggatctg 11220  
cgcgcacgcg gccacgcggt gttcgagatg gcgcccatt ccgtcgaggg ggccgagatg 11280  
gogcagcttg cgcgcgcgcg ccagacgtg ggcgacatca gcggcatcg cttcctggcc 11340  
ggcctgccc agcagcgcg ccgcacgggc gacgcggcc agcgcctcgc gcgccagacc 11400  
gcgcgctgc ccgcgctcgc gcgctgctg cgcgaatgca aggcgagtgc catgacacc 11460  
gcgggtgtgga tcgtgacctc gggcgcgcc tccagcctg cgcgcaacgt cccgcctt 11520  
gcctcgcgcc gccgcctcc cgatccggac gaggcgatcg tatgggggtt cgcgcgcacc 11580  
gcgatgaacg aattccccga gctgcgcctg cgcctcgtg actgcccga cccggagaac 11640  
ctcgagcgca acgcgcctc gctggtgcag gactgtctc cgccaccgc cgacgacgag 11700  
gtggtcctca ccggcgccgg ccgctacgcg ctgcgcctg gcgcggtgcc gccgcccga 11760  
tccgtgcgca ccgctacca gggcatcac cgcctcgact tcgcgcgcgc cggcccgctg 11820  
aagaacctcg cgtggcgcg cgacgtgcgc cgcaagccgc gcgcgaacga ggtggaaatc 11880  
gaggtgcgc ccgcgggct caacttcgc gacgtgatg acgcgatgg cctgctctcc 11940  
gacgaggcg tcgagggcg cttcgccggc gcctcgctcg gcatggagct ctcgggcgtc 12000  
gtcaccgcg tcggcaagga cgtgacctc gtcgcgcgc gcgacgaggt gctcgcgctc 12060



gcgcctcgg cggtcgccac gcacgtggtc accaccgccc attcggtcgc gaagaagccc 12120  
gcgggctgga cggtcgagtc cgcggccacc atcccgaccg cgttcttcac cgtgtactac 12180  
tcgttgaagc acctggccca gctgcgcgag ggcgagcgcg tcctgatcca cggcgccgcg 12240  
ggcggcgtgg gcatcgccgc gatccaggtc gccaaagtggc tgggcgcgga gatcttcgcc 12300  
accgcccggc cgcacgagaa gcgcgacttc gtgcgcctgc tgggcgcgga ccacgtgctc 12360  
gactcgcgca cgctcacgtt cgccgacgac gtgctgcgca tcaccaacgg cgaggcgctc 12420  
gacgtgggtgc tgaactcgct ctccggcgag gccattgcgc gcaacctgcg cgcgctgcgc 12480  
ccgttcggcc gcttcacgca gctcggcaag cgcgactact acgagaacac gcacatcggc 12540  
ctgcgcccgt tccgcaacaa catctcgta ttcggcgctg atgccgacca gctcatgaag 12600  
gagcgccccg acctcgcgcg ccagctcttc aacgagctga tgcagctctt cgagcagggc 12660  
gtgctctcgc cgctgccta ccgcgccttc gccgccaccg aggcggtgga agccttcgc 12720  
tacatgcagc actcgcgcca gatcggcaag gtggtgctgt cgttcgcca cggcggtgaag 12780  
gcgcagcccc cccggccccg cgtgccgcgc gagctggcgc tgcgcgcaa cgcacgtac 12840  
ctcgtcaccg gggcgctgtc gggcttcggc ctgcgcaccg cgcagtggtt ggtggacaag 12900  
ggtgccccgc acctgggtgt ggtctccaag agcggcgcgc agtcgctcga gaccaaggcc 12960  
gcgggtggcg acctcgaggg ccaggcgccc acggtgatcg cggaggcggt cgacatcacc 13020  
gaccgcgcct cgtgcaacg gctgctcgcc gaggtcgcg ccgcgctgcc gccgctgcgc 13080  
ggcgtgatcc acgcggcgcc cgtgatccag gacggcttca tcaccaacat gacggccgcg 13140  
cagatccgcg acgtgctggc gcccaaggtc ctcggcgcgc gccacctgga cgagctcacg 13200  
cgcggcgcca agctcgattt ctctgtgctg tactcgctcg cgacgacgct gttcggcaac 13260  
ccgggcccagg ccaactacat cgcggccaac tgcttcctcg aggcactcgc caaggcgcg 13320  
cgcgtcgagg gccagccggc gctgtgcgtg ggctggggcg cgatcgggga cgtgggctac 13380  
ctcgcgcgcc acgagaaggt gaaggaagcc ctgcgacgc acatggcgcg caccgcactc 13440  
gagtcggagg ccgcgctcgc ggtgctggag cagctcctgc tcgccgacgc ctcgggcctc 13500  
ggcgtgctcg acttcagctg gcgcaacctc cgccgcttcc tccccaccgc cgctcgcgc 13560  
cgcttcgcgc agctggggcg gcgcggcgac gacgcgaagg acgacgacga cggcgcgggc 13620  
gagctgcgcc gctcgccaa cgagctggac gcctccgagc tctcgcgat gttcaccgac 13680  
ctgctgcgcc gcgaggtggg cgagatcctg cgcacccgc cggaccgcct cgatacgcgc 13740  
cagcccttgc aggaatggg catggattcg ctgatggcg tggagctgct gaccgcggtg 13800  
gaggcgcgct tcggcgtaaa cctgccggtg atgtcgctct ccgagcagcc gtcgatcgag 13860  
aagctggctg accgcatcgt gcgcgcgctg aaggaccgga acgcccgcac cgaggccgaa 13920  
tcgcgcagcg accagatcga gcgcgtcgcc gcgcagtag cgcccgagct cgactcgcgc 13980  
caggtcgagg agctcacgca ggcggtcgat gacgcgcaag ccacgcagca gggacgtccg 14040  
tgaaggcgcg caacctccgc ggcctctcgc aggcagtga ggagcagctg atccagcgcg 14100  
tgctcgagca gcgcgtgcgc cgggtcgaga aggacaagcc ctcggcggtc gactccgcca 14160  
tcaacgcctt cgcggccgac ccgatgccc tgggcccacat ctcgagcgcg tactgccgct 14220  
tcgacatgca cccggtctac cagcagatgc agctggtgaa ggaagggtgc gcgaagctgg 14280  
gcattcccga tccgtacttc cgcgcgcacg acggcggtgg cgcggatacc acgcgcattg 14340  
ccgggcgcga gtacatcaac ttctccagct acaactacct cggtctctgc ggcgaccggg 14400  
ccgtggatgc cgcggccaag caggccatcg acgaatacgg gaccctcggt ctcgccagc 14460  
aggtcgtgt ccggcgagcg gccgctgcac cgcgagctgg agcgcgccat cgcgaggtc 14520  
tatggcggtc acgacgcccgt ggccttcac agcgggccac cgaccaacgt ctccaccatc 14580  
ggccacctgt tcgggcccgc cgacctcat gtgcacgacg cgttcgtgca caacagcatc 14640  
ctgatgggca tccagctctc gggcgccaag cgcattggct tcccgcaaa cgactggcg 14700  
gccctggacg agctgctgc gcgcgacgc gccacttc agcgcggtgt gatcgtgatc 14760  
gagggcgtct acagcatgga cggggactac cccgagctgc ccgaattctc gcgcctgcgc 14820  
cgccgccacc gcgcgttct gatggtggat gaagcgcaact ccttcggcggt gatgggccc 14880  
cgcggttcg gcatccgcga ccacttcggc atggccggcg acgaggccga catctggatg 14940

ggcacgctct ccaagacgct cgcctcgtgc ggcgggttca tcgccggcga gaaggcgtg 15000  
gtcgagcacc tgaagtctgc cgcctcgtgc ttctctaca gcgtcgcat ggccccgcc 15060  
gtggcgcccg cggcgctcgc cgcgtgaac cgcgtgcgc aggagccggc gcgctggcc 15120  
gccctgcagg cgcgcggacg cttcttctc gaggcgcgc gcgccggg catcgactg 15180  
ggcctctcgc agggcatggc cgtggtgccc gccatcaccg gcagctcgat ccgcgcggg 15240  
cgcctggccg aggcgatgtt ccagcgccgc atcaacgtcc agccgatcgt ctaccggcc 15300  
gtgcaggaga actcggcgcg cctgcgttctc ttctgagcg ccacgcacac cgaggagcag 15360  
ctgcgcttca cggctcgcga gctgcgcgac gcctggcgca agctctgagt ggccggcccc 15420  
cggctgcgcg tcgtcgtctg caccacggc ggctgaacg gcgcgtggt gctctcgcg 15480  
ctgctggccg cccccacgct cgaagtctcc gcgctggtga tctccagccg cgcgcgcgt 15540  
gcccatgagt cgtatggccg tgcgcgcgc ggatacgtgc gcgcgagcg cgctgcctat 15600  
gcgctgtacc tctggtgcgc caccgcgctc gccgacctgc tgctgcgcg cactccgag 15660  
ggcccggtcg cgcgcacgc cctcgcgcgc ggcatccgc tctcgcac gccgcgcgtg 15720  
aacgatgcca ccgcgcgcgc cttcatcgcg gggcgggcg cggacctgat cgtctccgcg 15780  
ttcttcaacc agcacatcga cgcgcagctg gctgcgtcg cgcgcgtggc cgcggtcaac 15840  
atccaccgct cgcgcgtgc gcacttccgc ggctggatc cgggtgtctt cgcgcgcctg 15900  
cgcggtgccc agcgccacgg cgtgagcgtg catcgcatc aaccggctt cgataccggc 15960  
gcgctgctcg cccaggaaac cgacgtcgag gccccggca gcttcttgc cgccaccgcg 16020  
gcgctctatg accgcggggc ggcgctcctc gccggccgtg ccgcggcgct ggccgccgac 16080  
ccgcgcggaa ccccccagcc ggcggggggc tcctacgact cctggccac ccgcgccag 16140  
gtcgccgctc tcggcgcgcg cggcgggcg ctgctccgcg ccgcgcacct gtggcgctc 16200  
gcgcgcggg ggccggccgc ttctgtaata gaatcagcg ggtagcgcg actcacgcg 16260  
cgcttttccc cgacccccg agggccatga aacactggt gaagcaacac gcaatcttcg 16320  
cgctcaccgt cctgctgcgc accgtggccg cgatcctcta cttcggcctg atcgctccg 16380  
acgtctacat ctccgaatca cgcttcgtg tgaggagccc ccagcgccag gtgcagaccg 16440  
gcctggtggg cgccctgctc tcgggcaccg gcttctcgcg ctcccaggac gacacctact 16500  
cgggtgcacga cttcatcacc tcgcgcgac cgctggcgca gctggacaag aagctcgcg 16560  
tcgcgaagct ctacacggcc gccaacatcg acttcatcaa ccgcttcccg gggctcgact 16620  
gggacgacag cttcgaggcg ttccaccgct actaccagaa gaaggtcac atcgacttcg 16680  
acaccgcgtc ctcgatcac gtgctgcgc tgcgcgctt cgagaaggcc gactcgcgcg 16740  
gcatcaacga cctgctgctg cagatggcg agcgctggt gaacgagctg aacgagcgca 16800  
gccgccagga cctgatccgc ttgcgcagc ccgaggtctc gctcgccgag gacaaggta 16860  
aggatgccg gctggcgctc tcgccttcc gcagcaacca gtcggtgttc gagcccgacc 16920  
gccaggcctc gatccagctg caggcgctg ccaagctcca ggaggagctc atcgccaccg 16980  
aggggcagct ggcacagctg cgcaagctct cgcgcgaaa cccgcagatc ggcgcgctcg 17040  
agaacaagtc ggcgcgctg cgcgtggcg tggcgcgca gtcgcgcaag gtgaccggcg 17100  
gcagcggtc gttcagcgcg cgcgccccg cgttcgagcg cctcaccctc gagaagggt 17160  
tcgccgaccg ccagctgggc gttgccctca ccgcgtgga gaccgcgcgc agcgaggcg 17220  
agaggaagca gctctacctc gagcgcatcg tgcagccaa cctgcccagc gaatcggtcg 17280  
agccgcgcg catccgctcg atcttcaccg tgttcgtgct gggcctggtg gcctggggcg 17340  
tggtgagcct gctggtggcg agcgtgcgc agcacgtcga ctaggccgtg gagcagaacc 17400  
cttccctcct gcgcgcctc ggggtgcagc ggcgcgtgct ctacgcgtg ctgatgcgcg 17460  
aggatcatc gcgcttcggg cgcgacgacc tcggcggtg gtcggtggtg gtcgagccga 17520  
tgatcttcac gctcgcgctg acggcctgt ggaccgcgc cggcatgaac cagcgctcgt 17580  
cgctgccgat cgtggcctc gcggtcaccg gatactccgc ggtgctggtg tggcgcaact 17640  
gcgcctcgc cgctcgatg gccatcgagg ccaacaagg cctgctctc caccgcaacg 17700  
tgcgctgat cgacgtgttc gtgacgcga tcatcctcga gatcgcgggc gccacggcgt 17760  
cgctcaccgt gctgggcac ttcttcact ggatcggtg gatgccgat ccgatcgaca 17820

tgctcaaggt cgcgttcggc tggttcatgc tgtcgtgggt cggcgccagc ctggcgctcg 17880  
ccatcggcgc cggcacccgc tactcgctgg tggtcgagcg gctgtggccg ccgacggcct 17940  
aactgctgtt cccgctctcg ggcgcgcct tcatggtcga gtggctgccg gagaaattcc 18000  
agcagttcgt gctcttgctg ccgatgggtgc acggcaccga gctgctgcgc gacggctact 18060  
tcggcaacgc ggtgcgcacg cactacgacg tgggctacat ggcaacggtc tcgctgggccc 18120  
tgtcgtgctg cggcatgcac ctgggtgcga tggcgcccaa gcgctggag gccgcgtgat 18180  
ccgggtcgag aacgtgacca gatgtactca cgcgcctcgg cccagaacgt gctgacgcgt 18240  
cagcttcgag ctccggccgc ggcgcaacat cggcatcctc ggcaagaacg gcgccggcaa 18300  
gtccacgctg atccgcctga tcagcggcgc cgagagcccc accaccgggc gcatcgctgcg 18360  
cgagatgagc gtgtcctggc cgtcgcctt cggcggcgcg ttccagaccc acctcaccgg 18420  
gctggacaac ctgaagttcg totgcgcgt ctacggcgct gactaccgcg acaaggctcg 18480  
cttcgtcgag gacttcaccg agctcggcgt gtatttccgc gagccggtgt tccactactc 18540  
gcacggcatg atcacgcgc tcgccttcgc gctgtcgatg gcggtggagt tcgactgctt 18600  
cctgatcgac gaggcgatgg tgggtggcga cgcgcgcttc cagcagcgt gccacgtcga 18660  
gctctccac aagcgcaagg accgcgcctt catcctcgtc acccagcagc ccaaggtgat 18720  
caagctctat tgcgagagcg cctgcgtgct gcacgagggg cggctgctgc cttccccac 18780  
cgtggacgcc gcctacgagt tctacatgaa cgaggtcatg caggacctcg ccccgagggt 18840  
tgcctgaacg ccccgcgcg gccgctgcgc gtgtcctcgc agctgcggcc ggcgttcgac 18900  
ggccacgccg gcatcccga ggaggcacgc ctgctgttcc gcggcctgcg catgatcgag 18960  
ggcatcgacg tggagggcct gctgcagcac agcgccacg tgcctgccaa gggcctgcg 19020  
ccgcgcggcg gcggcgacct cgcaccggac cgccagctga accgcctgtc gcgcgtcgtg 19080  
gtctccacca agcaggagct caccaacgcg cacgtggcga cggcgctgat ggccattcgc 19140  
cagctgctgg gcgcgcgcga gacctctac cgcttcgacg cggcgcaact ccgcgacttc 19200  
gtctggcagg cgctcttcgc gcgcacgctg cacgcgcgg acttcgaact ggtcaccgc 19260  
gccggctttc gcgtcgcgcg cgtgccgtgg acgggcatgc accgctgcgc gctggtcacg 19320  
cgcaagctcg gctacacgct gcatccgcgc atcgacactt ccgatttcga cgtgctgatc 19380  
gccgagacgc cctaccgcgc gcgggtcgcg gacggcacgc gcctcgtggt gcgctaccac 19440  
gacgcgatcc cgctcctcat gcgcacacc atctccgaca tgcctacca ccaggcctcg 19500  
cactaccggg cgctgcgcgc caacgtggcc tcgggcgcgc acttcgcgtg cgtctccgaa 19560  
gccacgcgca aggacctgt ctcggtcttt ccgagggtcg aggcgcgctc cagcaccatc 19620  
cacaacatgg tctgcacca ctacttcgc gagaccgcg acgcggggcg catcgaggag 19680  
atcctgcgaa cgcgcgcgag cgagcgcac aagggcgcgc ccaacggagc cgccaccgcc 19740  
gtcgcacgat cgcctcctt cctcgcgcgc gccgcgcaac ccagcgacc cggctacctg 19800  
ctcgcgtgg cgacgatcga gccgcgcaag aaccacgcgc gactgatcgc cgcgtgggag 19860  
cacctgcgca cctcgcgctt cccgggcctg cgctgggtgg tggtcggcat gcccggtgg 19920  
ggctacgagg cgctggtagc caagttcaag ccgtggctcg cgcgcggcca gctcttcgtg 19980  
ctcgaggacg tgcccgcccc cgagctgcgc ctgctctact cgcacgcgcg tgccacggtg 20040  
tgcccgagct tcggcgaggg cttcgatttc tcggcggtgg aggcgatgcg ctgcggcagc 20100  
ccggtgatcg cctcggagat cgcgcgcac cgcgaggtgt accgcgacgc cgccgagtag 20160  
tgcagccct attcgtggc cgacctcgc gaggcgatcg ggcgcgtcat cgaccggcg 20220  
gcgacgggccc tgcgccaggc gctggtcacg cgcggcaccg aggtctcga ggggtacacg 20280  
cccgagcca tctgcgcga gtggcgcgag tacctgctcg gcacggtgcg cgcggaggcc 20340  
ccgtgagcga tccgaaggcc accgaccagg cgtgggagga gtggggccag cgcgaccct 20400  
acttcggcgt gatcaccaac ccgcgcttcc gcccggggca gatggacgag gaagcgaagc 20460  
gcgaattcct cgcgtccggc cgcgtccatg ccgactacgt gatgcgatg gtgcacgcgc 20520  
acatcgcgcc cgacttcgt ccgcgcacca tctcgtattt cggctgcggc gtgggcccgc 20580  
tcgtcatccc cttcgcgcgc caggccgagc aggtgaccgc cgccgacgtc tccccctcga 20640  
tgctggccga ggctgcgcgc aactgcgcgc agcagggcgt ggccaatgcg cgcctcgtgg 20700

tctcggacga ctcgctcacc gggctgccgg gcccttcga cctcgtgcac tcgttcacgc 20760  
tggtccagca catcgatccg gcgcgcgggc gcgacatctt ccgccgcctg ctgggcaccc 20820  
tggcgagcgg cggcggtgggc gcgttgcaact tcgtctatgc gaagcgatc tacgcggcga 20880  
cgtacggcgt ggcgccaccg cccgagccgc cgcgcgcgc gccgcgcgc cagcctccac 20940  
ccagccgggc ggagatcaag gccgcgagca aggcgcgggc cgcgctggcc acgcaggcac 21000  
gggcgcggga accggcaccc aatcccagacc ccgacatgca gatgaattcc taccgcggcg 21060  
cggagatgct gttcctcgtg caggaagccg gcgtcacgcg cttccacgtc gagttcaccg 21120  
accacggcgg cgagctgggt ctcttcctct tcttcgcga gccctaggcg tccttcgagg 21180  
cgaaggagag gccagcgcg gtgatcgtcg cggcggtgctg cggggcgacc gccgcgcgg 21240  
ccgcgacttc ggcagccacg cgcctcgaggt cgcgcgcga cccgtacagg cgcacgcgc 21300  
agtagtgac ccagtaggag tccggacagg ccgacgcgat tccagcgcg cgcctcgacct 21360  
cgggcccgcg gcgcccctgg tgcacttggg gcgaggccca ttgcaggttc acggagggct 21420  
cgagggggcag gacggcgagc agcgcctcgc tcaccgccag cgcgcgcgc tcctgcgcgt 21480  
cgtcggcgag gcgcatcagc gtgggcaggg actccggcgc aagcgcgtgc gcggccgcga 21540  
gcgaatcgcg cagcatcgca gcgtcgcgc gcgcgagcgc caggcgggag cgggtgtagt 21600  
gggtccagaa cgaaggcgcc gcggcatcga gcgcgcgcgc gacctcgcg cttgcgtcgc 21660  
ccccggagag ccgggacgcg gccacaaca ggttcgcctc cgcgcgtgc gccggcacgc 21720  
tccgcaacgg cgcagcacc gcggtggcga cctcgcgcga accctcgcgc gacgagaagg 21780  
ccgagcgcgc gaactgcac agcgattcgc cgagcgacga atgttcggcg cagagccgcg 21840  
cgggatccac cgcgcgcgag gcgcggggcg cagcagcgc cgcctcctcg tcaccgcga 21900  
gcgcgtgcag ccgcaggcga tggaaacgcg cccactcgc cgacatcgac gccgcccgc 21960  
gctccagcgc accctcgacg gcgcatcgg tgcgcgcgc cgcatgtgc aggatggcc 22020  
aggctgcctg cgcgcggggg aggttcgtgc cttgcgcgc gagcgggtcg aggaccgcga 22080  
gcgcgcgcgc ggcgtcgcgc gcgccactga agtcgttggc caggccgatc agcgcggcg 22140  
cgagctcggg ggcgcgcgcg aaggcggccg ccgcgtcgc cccaccgcc gcgaggtcg 22200  
cgcgcagcgc gtagaggcgc gcgcggtggt aatagccca gaacgggtcg gggctcgcga 22260  
tcagtgaatc gaacgagcgc tcgacctcgc cgcacagtt gccgaggtcc acctggtcg 22320  
cgccccacag ggtgcgcgcga tgcgcggcca tcggctcgtc ggggacgaag gcgcggtgga 22380  
tcagctcgac gtgcgcgaag cgcgcgtcgc gcaacgcgag ctccaggcg ttcagccgct 22440  
cgaggctcgc cttgtactcg cgataggcat cgcggtcgcg gtogtattcc tgcgcggcac 22500  
tgtcgtccat cagcgaggcc gtgagccgct ggtgcgcgtg aacgggtgcg tatgcgtagg 22560  
ccgcgcgacg gtcgtagggc agcgtggccg cgcgcagcga cccgaacgcg gcgcggtcgc 22620  
ggaatcccag cttgcccage acgcggggtg ccaggtcgtg ggcttccatc tcgcccgggc 22680  
tccagtcgcg cagcgcgggc acgttcgagt agagccaca cggcgtgcg tagagctgcg 22740  
ccaggtcgcg gtccggggcg cctgcgcct cgaagatcgc cttccactcc gcgagtcct 22800  
cgtggccggc ggagaagatc ggcacgtggt cgcgaagat cgcagcacg tagggccgc 22860  
cgcgcacctc gagaagctcg accagcgagc gcagcgcgc gtcggtttcg gccatcaccg 22920  
cgcagtactt gtgcgagcg cctggcgca cgtgcccgc gaggaagtgc ggcgccacgt 22980  
agcgcggctc ggcgcccttg tcgcccagg agggctccgtg gccatcagc gagacgccga 23040  
acacgaagcg cggccgcga tgcgtcgagc ccgcgatcac ggcatccatc acggccgaat 23100  
ccgcgatgcy ctcgaactgg gcgcggcgcat gggcgaacgc ctgcgcggcg tggaaacgcc 23160  
cgaagccgag cttcttgtag acgcggtcgc ggttgtagaa ccagcgggtg aacggatgga 23220  
tggccacgct gcggtagccg tggcgggcga gcatgcgcg caccgcatac acccggtcgc 23280  
gcaactgccc catgtagggc agcgcgcggc agccttcctc gccaatgcg atgcgggtga 23340  
ggatctcgaa ctcggtgtcg caggtgttcg cgcgatcac cttcacgcgc aggcggcccg 23400  
tggtgcggcc cggccgcgtc gcgcgagga acggcagcgc gtcgcgcgc agcggcacgc 23460  
cgaacggcg caggtcggcg agcgactcca tcacaggaa gatgatgtcg gggcgctccg 23520  
cggtgccgg gcgcgtcgc gagcgtcct ggcgtccag cagcgcctc agccctcgc 23580

gcagccgccc cgccgcgacc gggcgggcgga tgctcccggc gaccagcacg gcgtgcagga 23640  
cgaagcccag gcgcccgcag ttacgcttgt ggtcgggtgg cgaatacgtc atcgcgtcga 23700  
actccaccgg ccagcggcgc aagcgtccgg cgacgaaggc gacgccgagc gcggcgagca 23760  
cgagcagcca ctgccggcga tcgtcgtacg cgtcgtgcag gtcgatggcc tcggtgatct 23820  
cgaacgcggc ccacgccagc gccgcgagca ctgcggcagc ggcaagcgcg cgcaccgcca 23880  
cgccgggggc atgcaggtac gccggcgcgga acaccagcgc ctccgcgagcg ctccaggaaat 23940  
cgcccggggc gaccacgcgg ccgaagaacg cgagcaggcc tcggttcacc gcgtagagcg 24000  
cgcccgccag cgccacggtc cccaccgtcg cgagcgcccc gaccgcgacg aagagcagcg 24060  
catggaagat ggccgagcag ccggcggttc aggccaggcg ccgcgcgcgc gccacgggtt 24120  
ccacgtcata gccaccacg aggcgcgcgt cgagggcggc ctgcagcacc gcgtaggcga 24180  
gcacgcagac cagcgtgccg cccagcagca cgccgaggaa acccgccaaa ccagccgtgt 24240  
cgtcaggat catgggacca ccagggacgt cttccggcca gtctataacg ccccgcccca 24300  
tgaaaaacgg cgaggcccg gggcctcgcc gtggagggaa ccgcgcaacc tcagggttgc 24360  
gggggtgcga ccgaccagcc gccggggcg atcagggttg cgctgccgat ggtggcgagc 24420  
ccgttcacga ggcgcaccgt gacgctgccg tcggtgtggc ggaagaccag gtccttgccg 24480  
ctgtcgccgt tcaggtccag cagctgcgtc acgttccaac ccgtgcccg cggcagcgca 24540  
tcgcccgcgg tgatgatgga aacaccatcc atcaggcgga tgtgcacgcg gccgtcggtg 24600  
tggcggaaga ccaggtcgtc cttcgcgctg ccgttcaggt cggccaccag gttcaccgac 24660  
caccgctgc cgccggcag gagctggcg ctggcgccga aggtggtgcc gttcatcagg 24720  
aacagggtgc cgccggcgtc ggtgtggcg aagatcatgt ccgcgcgtcc gtcgcccgtg 24780  
aggtcgccca ggtggcttac cgtccagccg ctggcgcccg agaggagcc cgctccggcc 24840  
gtcaccgtgg tgccgtccat gatgtagatg tagccgcggc cgtcgtctgt catgaagacc 24900  
aggtccgcct tgccgtcgcc gttcatgtcg cccgcctggg tgagcgtcca gccggtcccc 24960  
gccccgaaga gctgggcgt gccgatgatc gtcgtgccgt ccatcaacca gatgtgcgcg 25020  
cgtccgtcgg tgtggcgag gatgatgtcc gccttgccgt cgcggttcag gtcgcccgtg 25080  
tggctcacga tccagccgag cccggcgggc agcagctcct tgccggccgt caccgtagtc 25140  
ccgttcacga tgtacgcgta ggcgcggcca tcggtgtggc ggaacaggat gtcggcgcg 25200  
aggtcgccat tgaggtcggc cgtctggctc accgtccagc ccgcgcggc gccgatcagg 25260  
ttggccgagc cgggtgatggc ggtgccgttc atcgtccaca ccgcgatgcg cccgtcgctg 25320  
ttctgcagca cgaggtcgt cctgccgtcg ccgtcaggt ccgagacgac cggcttggtg 25380  
ggcggcgtga cgccaccgga aaccggcagt gccagccgt cgatccaggc gcggtcgcg 25440  
cagttggcg catccggggc cggcggttc caggtcgcg agttgcaagg cgtggagagg 25500  
aagttcgaga aacgccacac cagcgtgtgc gcgcccgcg tcaccgggaa cgacaccagc 25560  
ttccagccc aggcctggt gccgcgtcg ctgaacacca ccgtgccgtc gatcaggaa 25620  
tcgaacttgc cggcggttgg gaagctcgac acgcgatagg cgaacgcgac gttgcccgcg 25680  
agcagcgtgc ccgcgaacga gaggtcgag ttaccgtcg tcgagttggt cgggtcgctc 25740  
gtcaccacct gcgccgaac aaggctggtc gcgcctcga aggcctggtc ggagccgacc 25800  
gtccatcggt tggcgccg ccaggtcgtg aagccggcag gcagcgtgcc acccgtggg 25860  
cacggcgaga ggatgatgcc ggtcgcggtc gccgggccc cgatgaagac acccgtggaa 25920  
cccgtcgcat tggagagcgc gacgttgaa gtctcgttgc cttaccgat accgtcgctc 25980  
gccaccggga ccgtgatggt cttgggcgtg gtctcgccgt tggcccagct gagcgagccc 26040  
gaggtcgcg tgtagtcggc gccgaagtg gcggtgccgt tggaggtggc gtagctgacc 26100  
gagatggcac ccgccgagcc gccgatgogg ctacgggtga gcgtcacgt gccggcggtt 26160  
tcgcggccg cgaagggtgt gccggtgaac tgcaccgaac cgggaacggc gggcggggtg 26220  
accgtggact tgagcgccga gagtgccgcg cggttggtgt tcagcgagag cgcgtcggtg 26280  
gcctgcgcgg tgccgcacgg ttgcggcgca ccgacacccg tcgtgcacga gagggccggg 26340  
ttggagaagc ggtagacct ggtcgctgag tcatgcagca gcgacatgat ggtggtgaag 26400  
ttcgcccgct cgttcgctcg gcaactcggg cgctggctca cggcgaggc gccgcggccg 26460

gagccgttgg gcagggccga gttgcaggtg aggttgccgc tcgcgcaata gaagtggccg 26520  
tacgagtacg cgaacgcacc ttgcggcggg ctcgcggtgc cgcccgttc ccaggcctgc 26580  
gagggcggt cgtggcggtt gcccatggca tggccgagct cgtggatgaa gaccactcg 26640  
caaccgaaca ggcagccgga agcaacggcg taggcaaagt tggggtccgg cgtgctggtg 26700  
ccgatccagg ctgcgccgtt gccgccgaag tcggcgccgt tgcgcatgaa ggccaccatg 26760  
tccgcgccgt attgcgtgcg gatggcctcg atgttgccga acgtcgccgc gtcgaagctc 26820  
gcgtcgcccg gggatgatgg atgcatcgcc gtggtgtcgt cgatggcgtc gtcgtagtgc 26880  
acctgcgtcg cattcaccag gcgcagcgtg atggcgacct cgctgtcggc atacgcagtg 26940  
ttggcgcgcg tgacgaggaa gttcaggcgc gtcatcaggc tgccgccgag gcgctggccg 27000  
aagccgcgcg agtacacgat catcaggtcg atgacgttct gcggcggtgg cgtcacctgg 27060  
gcgagccgac ggtattcacg ccggggatgg cgatcagggt ctcggggctg ccggtcgca 27120  
cgacgcgcga cttcgagccg gtgctggcca cggtgtggtc ggccgccgag ctgatgatcg 27180  
gcatgctggc ctgctcgcc gtcatgtcga ccaggtagtc gacggtgccg cccggaatca 27240  
ggcggtactc accctgcggg gtcgagaaca cgccgaagga gccttcaggc ccggtcgtga 27300  
cgatcgcgcg gtggttgatg ccgtcggtct tgcctctcgc gatccacgac gtgatgccgt 27360  
cgccgtgggc ctggagcaac tcgaagacat acgagtggcg caccctgtt gggagcgaca 27420  
gctcgacctc ggagcgcgga ctcagggcct ggaggccgc ggcggtgaag cgcaccgctt 27480  
gccgggcaac ggcgctgcc ggcccgga tcgcggcggc ggtgtcgag accaggatct 27540  
ggggaaccgc ggcgaaggcc acggaagaga agacgatcgc cagcgcccg aggaaggcg 27600  
ctgtgatccg ttgcgtgaat gcgttcatgt ggagctccgg aagttgacct atgccaatc 27660  
cgctatgtcg cggagatgtg gacaaaagg atcgagcgg cgtagcacc cgccccgga 27720  
gggatgctcc aaaaggacta cgagggtgct acggctgggt tgggtggcga ggccgaagga 27780  
gaactccttg tggctctccg gcgacttaag gttgcgagg aaccaccgac caccaccgg 27840  
ggccgatcaa gtccgcgtg ccgatcgtg tgacgccgtt catcaggcga atcgctacgc 27900  
gcccatcgtt gtggtggaag accacgtcca tcgcccgtc gccgttgaag tcgaggagct 27960  
gcgtcaccgt ccagcccgcg aacggcgga ggatgtcggc ggccgccgag atcgcggtgc 28020  
cgttcatcag gcgcacgtgc gcgcggccgt cggtgtgcgc gaacacgagg tccgcccg 28080  
cgtcgccgtt gagatctccc accaggttca ccgaccatcc ggtgcccgcg ggcaacagct 28140  
cggcgctggc gccgaacgtc gtgccatcca tcaggaacag gtgtgcgcgg ccgtcgttgt 28200  
ggcggaagac gaggtcgcc ttctggctgc cgctgaagtc cgcgacgtg ctcaccgccc 28260  
agccgctgcc gggcgcgagg aagcccacgc cgccggtgat cgccgtgccg ttcattgatgt 28320  
atgcgtagcc gcgccatcg acgttgagga agacgatgtc ggcccttgccg tcgcggttca 28380  
tgtcgccgtt gccgaccacg ttccatcccg agcccgcggc cagcaggctc gcaactgccg 28440  
tgatcgcggt gccattcatc agccagatgt gggcgcgctc gtccgtgtgc tgagcagcag 28500  
gtcgcccttg ccgtcgccgt tcaggctccg cgtgtggctg atggtccagc ccgtgccggc 28560  
ggggaacagc tccttgccgc ccacgaccgt gagcccttc atctggtacg tgtagatgcg 28620  
cccgtcggtg tgctggaagt agatgtcggc ctggctgtcg ccgttggaat ccgcggccat 28680  
tgatcacgct ccagccggcg ccggcgggaa tcaggctcgc cgaaccagt atcgctcgtt 28740  
cgttcatggt ccacgcggcg atgcgcccat cggtgtgctg gaacaggaag tcgctctgcc 28800  
cgtcgccgtt caggtcggtc gcccatggg tcgggaacgc caccttcgcg aggaaggcat 28860  
cggagaaacc cgtgagcaac gtcttgtagc cgccggcggt ggtcggatag ttcccgacg 28920  
acgtgcggcc ggccacatag gccgcgcccc gcgcacgac gccgacgcc taccacact 28980  
cgacgttggt gccgcgatc agcgtcgaga acacgatgtc gccgttgccg acgcgagtga 29040  
ggaacgcgtc ggcacgcgc ccggggccct gcaccgggt cagggtcggg aagccggcga 29100  
cctgcgtgta gccacgacg agcgcacgc cattcgccgc cagggcgacg tccaggggat 29160  
tgtcgtaggc cgagccgccg aagatgccc acgattcgag cgcgcgggtg gccgagaagc 29220  
gcgtcacgaa cgcacgacg gtgccggcga aattgcgcac cgttcgggtc tgcgggaagc 29280  
tgggcaggcc ggtatcgccg acgacggtcg cctggccgtg gaccgggttc accgcgatcg 29340

cggtcacgtt gtcgccgctg ccgccgagga acgtggaata cgcgaccggg ccgccgctcg 29400  
caccgatctt ggtgacgaag ccgccaccgg gttgggtctg gtacgcattc acgttgccgga 29460  
acgtcgagga gcgcgtggag ccgcgcagct acgtgttgcc cgcattgtcc gttgcgatac 29520  
cgcggccgat gtcatcaccc ctgccgccga gatacgtgga atagacgaag ccgccgttga 29580  
acgagagctg gtgacgaacg cgtcgcgatt gagggccggtg ccgacgaaca cggctgcagc 29640  
gcgctgccgc cggtcgggaa gccgcgcgcg gccgaaccgg tgacatgcat gaagccggtg 29700  
ttgtccacgg cgatcgccat gccctcgctg tcgccggagc cgcgcagcag ccgcgagaag 29760  
atgacgttgc cagtggattt cagcgcgctg acgaatgcat cgcgcccggc ggcatggttg 29820  
gtgcaatcgg tgacggcgaa gcaatgcggg aacgcgttgg tgccggtgga accggtgatc 29880  
gcgacgaact cgttgccgcc ggaggcgccg agcgcagcc cgcggagata cgcttcggcc 29940  
ccgagcgtgg tcgagtacac cacctgctgc gtggtggggt tgaacttcac caccgtggct 30000  
tcgtacgtgc ctgcggtata ggtcgcgagc ccgacgtaca cgttgccctg cgcgtccacc 30060  
ttgaccggcg tggtcactg ctgctgcgcg gcgccgccca ggtacgtgga gaacgccatc 30120  
accgggtcga tcaccagcgt gcgggtgctg tcgtagtcgg cgatcacgaa gccggcctcg 30180  
gcgccctgct cgcgccagaa gagctcgaag cgcgcggcca ccggcacgcg cgaatcgccg 30240  
acctgctgga aggtaccgg cgcgtgctgc gtaaatcct cctggcccac gcgcatgcgc 30300  
aggttgccct cgccatcgat ccacgcgtgg tcggcggacg agaggctgag gcggtctg 30360  
cgcggatcgg cgcgcggggc gaccacgaag tcgtattcga gcgtgccctc cttgccgtac 30420  
acggtgaggt ccacgcgcgc atacaggtcc ttcagcgaca cgcggccgaa gtgcggcacg 30480  
tttctgcgcc ggccggcggc cgtggtgccg ctgtaatagt ggctgagggg ggcccggggc 30540  
tcctcggcct cgatgaccgg atcggaagcc ccgcgaagc gcacgcgcaa cagcgaggcg 30600  
ctggcagctt tcctggagcg tggctcgaag gcaatgccat cgcggctcac ggagacgcgt 30660  
ccggcctgcc cgcgcgagac gtagagcgca tcggggccga actggccctc gttgcgctcg 30720  
aacgtgatgg gcgcgttcct gagggcatca gggacgacgg catggacggc ggccgggagt 30780  
gcgagggcgg ccgcgacgag cacggccagg gggcggagcg ccgggcgggc gccgctggag 30840  
ctgcgggtca tggggcagtg acctttctgt tgttatggct gcttgttgtc tttgcgggat 30900  
ccacagctcg gatccagctg gcgatactac cagcgagcga gtggttttcc gtgaccagct 30960  
cccgcagtcg ggccccggcg acaacggcgg catcgcggtc ctttgccagg gcacgcacgg 31020  
cctcgatcca gtccgcgtcg cgggcccagca cggcgggggc atcgcggtag ggctcgagat 31080  
ccgaggccac caccgggatg ccgaggcagc cgtactcgag cagcttcagg ttgctcttcg 31140  
cgcgattgaa cgggttgctg cgcagcggtg ccaccgcgac gtcgagtgcc agcgacgoga 31200  
gcttctccgg gtactgcgcg atgggaccca tgcgtgcac ttccgcggcg aacggcgcca 31260  
gctcgggctg gcacaggccg aggaacaccc agtcgatctc gcgatgcgtg gcgcgcacca 31320  
cgggctcgag caacttcagg tcttccccat gctgtttcgc ccccgcccag cccacgcgcg 31380  
gacgcgcgcc acccgcgga cggttcgcca ggccggccca acgctccgca tcgatcgctg 31440  
tcgggatcac gcgcacgtcc ttcgcgcgcg ggccgaaggc ctccgcgagc ggtgcgggtg 31500  
acaccacgag gcggtcgcac aacgccaccg cgcgcgcgat gcgctgggcg atgtccgggt 31560  
agatcgctcg cgcatacgga ttgccggggc gcagttgcgt gagcaggtca tccaggccca 31620  
ggaccttcag cgcgcgtccg tggcgcgcga gcacctcgag cgaggtgagc tggtagttgt 31680  
ggaagaagtt gtgcgcgagc acggcgctcg cgtcgaggcg ctgccactcg acgcggttg 31740  
gcgcgcagcc cgtggcgctg tcgccatca tcacgacctc gacgcggccc gcgcgtcca 31800  
gcgcgcgcga cggctggcgc acgcgcacct cgcggagcc ccagcggtcg aacggaaacg 31860  
cgcagagccg aacgcgtgc gtgtcgttgc cgcgcggcgc gaagcgttcc acgggcccgt 31920  
agccctcgcc gcccatgcgc aacgcgggt ggtagtgcgg atcgctcgct agcaccggct 31980  
gccagcgctc gcgcatccac gcctcttcgg aaggaatgac gagcttcgcc accggaacga 32040  
cggcatcggc gcgcgcatcg gcgagcaacg ggaagtccgc cgcgatgccg gggcgcgcaa 32100  
ggatgtcgaa gcccgctcg cgcagcccga ggcacaggtg cgcctcgcg aacggccccg 32160  
cgcgctcgat ctcgctgcagc gcctgcgcgg agagcgccgc gtgcggttg accagcgcca 32220

gacggccgcg cgcgacgctc acctcgcgcg gcgatgcgta gagtgcgcg aggcgctcct 32280  
cgccctgcgc gcgcggcgga gcgcccagaga cggcccacgg cccgcgcgcc gcgatttccc 32340  
agccgggaac gcgcacaccg ccgggggccc gcaggtcggg cgcgaggga cccgtggagg 32400  
ggcctgcgat gccctgctcg aggcgctcga gccagcgggc gcggaattcg ctgcaacgcg 32460  
catcgacgat cgcgatccac agcgtcgccg cgttgcggat ggcgcgcgcg aacgcggagc 32520  
cggtttcgcc gccaccgga accaggtgcg tgcgcacgcc gcgcccgcgc agctcaccgg 32580  
gccagcgttt cggcgccctt cgcgctcgaga tcgacgtaca cgtggccgat acgttcgcgg 32640  
cagcccttca cgagcgctc gatgcacgcc accagccgct ccggcgcgctg cgtcgcgctca 32700  
atgaaggcgc aaatcgtgcc ggccccgacg gcgcgcgcgc acgatgcggg cgcacgcgac 32760  
gcagctggca ggaacgcggc ctgcgtgcgc gtgcggcgga agtgctccgc gaccacgcgc 32820  
tggtggcgct cgcgcgcggc gtcgctttcc tgcgcggcga tgttcagggtc gaggcgatgc 32880  
acgagcagct cgcgcagtg caccacggcg ctgcacggc cctcctccgc gaggcgagc 32940  
gcgagatcgg tgacgcccgc ccagccgccc gcgcgcagcc cgccacgctc cagcaacgca 33000  
gcgcggcgca gcaccgcgag cccggcgagc ggctggtgtc gcgcgaactc cgcgccgaac 33060  
gtggggacga acgcccgatc gtagcgcttc ccgtgcgcgc tcagggcacg actgcagccg 33120  
tagagaagct cggcgcttcgg gttgcgtgcg accgcaccgc ccaccgcgct cggcgcgcca 33180  
tcggccagcc ggtcgccctc ctccaccagc acgacccatt cgcccttcgc gtccgcgacg 33240  
gcgcggttga cgcgcccag cgccgcgtcg ctgcgcgcgc gctcgaggac caggtctcc 33300  
agggcgatcg cggcagtgct cctggccggc aaggcctcgc gcgccatcca gtggtcgtag 33360  
gcacgcgcgc ccagcgcggt gcgctcgcg cggcgtggaa agccgaggca cgccgcggcg 33420  
aacaacggat cgatcggttc gggcacagca tccaccgcgc catcgcgctc cgccgccagc 33480  
tcgcgataga gcgccaggta ctcgccgca ttcgcctcga tcgtcttcgg gcggtggtac 33540  
ttcaggttct ccgcgacgcg atcgagcgcc gtgcggtcgt cgcgcaggcg cagcaggagg 33600  
gccacgatcg cgccgggatc gttgtgcggc agcaggaagc cgtctcgcgc atcgcgacg 33660  
cgctcgggaa tggcgccatc gcgcgtgccc gccaccggga tgccgaggcg ctgtgcctcg 33720  
ctcaacacca ggcagaaggc ctccctccatc gtcgagggca gcaggaccag gtcgatgccg 33780  
cgaaggatgc gcggcaggtc cgtcgcgga taggcgcgcg gctgcgatg gcccgccgca 33840  
ttcgccgcgc cttcgagcgc gagatcgcg ggacccagg cctcgatcgc gatgcgctgc 33900  
cccgccagcc ggcgcgcggc ctcgatgatc aggtgcgcgc ccttgcgcgc gccgaagcgg 33960  
ccgaggaagg ccacgcgcag caccggccgc gcttcggggc gcggaagcag cggcaggctc 34020  
gcaatgccgt gcccgatgac gcgcagcttc gcttcgcagc tcgcgccgaa gcccttgcgc 34080  
atgcgcatcg ccgcgtactc ggacgggcag aggatcacgt cggccgcgctc cagggcatcg 34140  
gacgccgcgg cgaagcgcg ttcgaggtag ccggtcacgg acgccggcg cgtcgcccc 34200  
acgatgcct cgctgcggcc gcgcaggcaa tccacgcac cgccatcggc gccgcgcgcg 34260  
gcggcgcggc cgacggcag gtcctgcggc ccggtcatca tgttgtagtc ggcgcacagg 34320  
tacgagaagt cgtgcgcgct gatgaccag cgtgccccgc actcacgcgc gatgcgcggc 34380  
agggcgaggc tgttccagcc caccaacgac tggaagtgca cgacgtcgta gccccggac 34440  
agcagttcgc ggaaggccgc ctccacgccc gcgtcgga cgcgcgcctc gtgcccgatc 34500  
acgcggatgc cggcgctcat gcgctcgggg ttcaggcgcg cgagtcgcac gtgcgtcgcc 34560  
agggcgtcgg tgggtgaagtc ggtccacgcg cgcgggccct cgggcagcac cagcgtcgag 34620  
gcgacgtcct tgcgcagcag gttcaccagc gcgcgcgtgt gtcctcaat gccgcgcgcg 34680  
gagtcgaggc gatgcagcac ctgcagcacg cgcgggcggc ccggcatccg ctgcgcgtcg 34740  
gccgcggcgt tgacgcgctc ggcgggcgcg cgacggggt tcgcgatcgc ccacgcgcgc 34800  
acgcccgcga cgtaggtggg ccagcgcgcg tcgaggcgca cgcggttggc ctgcgcgcgcg 34860  
gcgtcgatgc cggccacccc gccgaacgag gcctcgcccc cgtggtgcac gtagacgtcg 34920  
tcgcagcacg cgatctgcca gtcccggctc agcgcgcgca tcgagaggtc gtttctcctcg 34980  
ccgtagccgc ggccatagag cgagtcgaac gggcgcgatt gcaccacac ctgcgcgcgcg 35040  
accagcatgc agaagccgac cgccgtcggc aggcgcggat actcgcgcg gctggcgcgcg 35100



gcgacgcagg cggagatggc atcgacgtcc ggctcgccgt cgggccgctc gaagatatcg 35160  
gcgacgccgg gcaggctgag gatcgtcgcg cggttggaga gcgggcacac gatgccgacc 35220  
cgcgcacgcg tgctcgcgca acgcagcagg cctcgagcc aaccggggcc gacctgcgta 35280  
tccgaattga gcaccacgaa gtgcgcggcg ctgcgggcca ggcccgcagc gaccgcgcct 35340  
gcgaagccga ggttcttcgc attggtgacg acacgagtgc cggcacgacg cgccgcgaag 35400  
ctgcccagca acccggcgag gcgcggatcg gtgctcgcgt catcgacgag gacgatgccg 35460  
tgccgcggcg ccgtgtggcg ggccagcgac gcgaggcagc gttcgggtgc gtcgtggccg 35520  
ttgaagaccg ggacgacgac tagcgccggc gccatccgac cggcgcttcg agctgggcca 35580  
atcgcgccag caggcgtggt tgctcctgcg agagtgcctc gagctgcttc tgcgaggcga 35640  
cgtaggcgcc tccatgcgct cccaggtgcg tgcgcgcctc ctcgagctgc cgcagcgact 35700  
gctcgtggcg accggagacg tcgtgggtcg ggtcggagag cagcgagaag cgcccggccc 35760  
ccgcgcgtcac cgcctcgcgc tcgcgcgaca gggcgaggaa gtacatcgca tcggcgacgc 35820  
ccgcgcggg tttcgcagcc tccggcgccg aagcttgacg caactgcacg gcgtcgggca 35880  
ccgcgtgcag cggccagatc gccgaatacg catcgacgcg ttgccgaac agggccacct 35940  
ccgggaaccg cgccttcagc acggcgagga attcgtgctc gtagagctcg cgcacgtgat 36000  
gcgggttgcg gtagtcgcgc tggtcgaggt acacctcgcg gttcggcgctc gagaccagca 36060  
gcaggccgcc gggcgccagc acgcgcttcg cttcgtcgag caggcgctcg ggatcgggaa 36120  
tgtgctcgag cgtctcgaag gagacgagga gatcgatgct tgcgtcatcg cacggcagcg 36180  
actcgacgcg gccctcgacg tactgcaggt tctgcgcggc accatagcgg cggcgcgctc 36240  
gcgcgatggt ctcgcgggcg acatccgcgc ccaccacgga cttcgcgcgc gtcgccagca 36300  
gcgcgagacc gtagccctcc ccgcaggcga tgcgaggac gcggcagcct ccggcgagcg 36360  
gcaacgcaaa gtggtagcgg tgccagtgtc cgtaccagat ctgcgccctc aagccgggct 36420  
ggaaacgttc gtgttccatc gggagagtat aggtggatgt gcaaccgccc tccggagagg 36480  
gcggttgca tgatcttatt cgagatcggc tgctacggaa gcaccggcaa cggctgcggc 36540  
ggcgccaccg accatccggt gccgcgccc agcagggttc cgtgccgag cgtggtgagg 36600  
ccgtccatca ggcgcaccgt gatgcggcca tcgacattct tgaacaccat gtcgagcttg 36660  
ccgtcgccat cgaagtcgag cagttgcgtg acgggtccagc ccgccccgc cggcagcacg 36720  
tccgcgcgc tgaggatggc ggtgccgttc atcaggcgca cgtgcgcacg gccgtcgggtg 36780  
tgccggaaca cgatgtcacc gcgcccgtcg cggttcatgt cggccacgtg gctcacgac 36840  
cagccggtgc ccgcgccgag gagctcgga ccggcgccga acgcggtgcc gttcatgatg 36900  
aagagatgcg cgcggccatc ggtgtggcg aagaacatgt ccgccttgcc gtcgcgcctc 36960  
acgtcgccca ggtgcgtcac cgtccagccg ctggcgggcg agagggaagc cgcgcggcg 37020  
gtgatggtgg tgccgttcac caggtagatg tagccgcggc cgtcagcgtg gatgaagacg 37080  
atgtcgtccc tgccgtcgcc gttgaggtcc ccggtggcca cgactctcca gccgtcgcc 37140  
ggccccagca gctgctggct gccgatgatg gccgtgccat ccatcagcca cagggtgcgag 37200  
cggcgctcg cgttcgcgag caggagatcc gccttgccgt cgcggttcac gtcggcggtg 37260  
cggtcgatgc tccagccgag gccggcgccc agcagctcct tgccgcgggt caccgtgagg 37320  
ccgttcacg tgtacacgta cacgcgcccg tcggtgtggt gggatcctct agagtcgacc 37380  
tgcaggcatg caagcttgag tattctatag tctcacctaa atagcttggc gtaatcatgg 37440  
tcatagctgt ttcctgtgtg aaattgttat ccgctcaca ttcacacaac atacgagccg 37500  
gaagcat 37507

&lt;210&gt; 3

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Sense primer.

<400> 3  
ggscckcss tsdcsrtsga yacsgc

26

<210> 4  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Antisense  
primer.

<400> 4  
gcbbsryyt cdatsggrtc scc

23

<210> 5  
<211> 773  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: OriT derived  
from plasmid RP4.

<400> 5  
gatctgtgat gtacttcacc agctccgcga agtcgctctt cttgattgga gcgcatgggg 60  
acgtgcttgg caatcacgcg ccccccccg cggttttagc ggctaaaaaa gtcattggctc 120  
tgccctcggg cggaccacgc ccatcatgac cttgccaagc tcgtcctgct tctcttcgat 180  
cttcgccagc agggcgagga tcgtggcatc accgaaccgc gccgtgcgcg ggtcgtcggg 240  
gagccagagt ttcagcaggc cgcccaggcg gcccaggtcg ccattgatgc gggccagctc 300  
gcggacgtgc tcatagtcca cgacgcccgt gatctttagc ccctggccga cggccagcag 360  
gtaggccgac aggtcatgc cggccgcccgc cgccttttcc tcaatcgctc ttogttcgtc 420  
tggaaggcag tacaccttga taggtgggct gcccttcctg gttggcttgg tttcatcagc 480  
catccgcttg ccctcatctg ttacgccggc ggtagccggc cagcctcgca gagcaggatt 540  
cccgttgagc accgccaggt gcgaataagg gacagtgaag aaggaacacc cgctcgcggg 600  
tgggcctaact tcacctatcc tgcccggtcg acgcccgttg atacaccaag gaaagtctac 660  
acgaaccctt tggcaaaatc ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat 720  
cgctataatg accccgaagc agggttatgc agcggaaaag atccgtcgga tct 773

<210> 6  
<211> 2012  
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Psi C31  
integrase.

<400> 6

```
agatctcccg tactgacgga cacaccgaag ccccgggcgc aaccctcagc ggatgccccg 60
gggcttcacg ttttcccagg tcagaagcgg ttttcgggag tagtgcccca actggggtaa 120
cctttgagtt ctctcagttg ggggcgtagg gtcgccgaca tgacacaagg ggttgtgacc 180
ggggtggaca cgtacgcggg tgcttacgac cgtcagtcgc gcgagcgcga gaattcgagc 240
gcagcaagcc cagcgacaca gcgtagcgcc aacgaagaca aggcgggcca ctttcagcgc 300
gaagtcgagc gcgacggggg ccggttcagg ttcgtcgggc atttcagcga agcgccgggc 360
acgtcggcgt tcgggacggc ggagcgcccg gagttcgaac gcatcctgaa cgaatgcgcg 420
gccgggcggc tcaacatgat cattgtctat gacgtgtcgc gcttctcgcg cctgaaggtc 480
atggacgcga ttccgattgt ctcggaattg ctcgccctgg gcgtgacgat tgtttccact 540
caggaaggcg tcttccggca gggaaacgtc atggacctga ttcacctgat tatgcggttc 600
gacgcgtcgc acaaagaatc ttcgctgaag tcggcgaaga ttctcgacac gaagaacctt 660
cagcgcgaat tgggcgggta cgtcggcggg aaggcgcctt acggcttcga gcttgtttcg 720
gagacgaagg agatcacgcg caacggccga atggtcaatg tcgtcatcaa caagcttgcg 780
cactcgacca ctccccttac cggacccttc gagttcgagc ccgacgtaat ccggtggtgg 840
tggcgtgaga tcaagacgca caaacacctt cccttcaagc cgggcagtca agccgccatt 900
caccggggca gcatcacggg gctttgtaag cgcatggacg ctgacgccgt gccgacccgg 960
ggcgagacga ttgggaagaa gaccgcttca agcgcttggg acccggaac cgttatgcga 1020
atccttcggg acccgcgat tgcgggcttc gccgtgagg tgatctacaa gaagaagccg 1080
gacggcacgc cgaccacgaa gattgagggg taccgcattc agcgcgaccc gatcacgctc 1140
cggccggtcg agcttgattg cggaccgatc atcgagcccg ctgagtggta tgagcttcag 1200
gcgtggttgg acggcagggg gcgcggcaag gggctttccc gggggcaagc cattctgtcc 1260
gccatggaca agctgtactg cgagtgtggc gccgtcatga cttcgaagcg cggggaagaa 1320
tcgatcaagg actcttaccg ctgccgtcgc cggaaggtgg tcgacccgtc cgcacctggg 1380
cagcacgaag gcacgtgcaa cgtcagcatg gcggcactcg acaagttcgt tgcggaacgc 1440
atcttcaaca agatcaggca cgccgaaggc gacgaagaga cgttggcgct tctgtgggaa 1500
gccgcccagc gcttcggcaa gctcactgag gcgcctgaga agagcggcga acgggcgaac 1560
cttgttgcgg agcgcgccga cgccctgaac gcccttgaag agctgtacga agaccgcgcg 1620
gcaggcgcggt acgacggacc cgttggcagg aagcaattcc ggaagcaaca ggcagcgctg 1680
acgctccggc agcaaggggc ggaagagcgg cttgccgaac ttgaagccgc cgaagccccg 1740
aagcttcccc ttgaccaatg gttccccgaa gacgcgcagc ctgacccgac cggccctaag 1800
tcgtggtggg ggcgcgcgct agtagacgac aagcgcgtgt tcgtcgggct cttcgtagac 1860
aagatcgttg tcacgaagtc gactacgggc agggggcagg gaacgccccat cgagaagcgc 1920
gcttcgatca cgtgggcgaa gccgccgacc gacgacgacg aagacgacgc ccaggacggc 1980
acggaagacg tagcggcgta gcgagacacc cg 2012
```

<210> 7

<211> 84

<212> DNA

<213> Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: Left arm of  
pPAOI6 transposon.

&lt;400&gt; 7

ctgtctctta tacacatctc aaccatcatc gatgaatttt ctcggtgtt ctgcatatt 60  
ggctcgaatt cgagctcggg accc 84

&lt;210&gt; 8

&lt;211&gt; 94

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: Righth arm of  
pPAOI6 transposon.

&lt;400&gt; 8

gatcctctag agtcgacctg caggcatgca agcttgccaa cgactacgca ctagccaaca 60  
agagcttcag gggtgagatg tgtataagag acag 94

&lt;210&gt; 9

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Primer AmF.

&lt;400&gt; 9

ccctaagatc tggttcatgt gcagctccat c

31

&lt;210&gt; 10

&lt;211&gt; 34

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Primer AmR.

&lt;400&gt; 10

tagtaccgg ggcaccaacg tcattctcgtt ctcc

34

&lt;210&gt; 11

<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer OriTF.

<400> 11 32  
gcggtagatc tgtgatgtac ttcaccagct cc

<210> 12  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer OriTR.

<400> 12 37  
tagtaccg g gatccgacg gatcttttcc gctgcat

<210> 13  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer Fint.

<400> 13 32  
aacaagatc tcccgtagtg acggacacac cg

<210> 14  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer PJ.

<400> 14 21  
cgggtgtctc gcatcgccgc t

<210> 15

<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer  
amyE-BamHI

<400> 15  
atcgcaggat cctgaggact ctogaacccg

30

<210> 16  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Primer  
amyE-EcoRI

<400> 16  
cgactgaatt cagatctagc gtgtaaattc cgtctgc

37

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/07765

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/10 C07K14/00 C07H21/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, BIOSIS, MEDLINE, EMBASE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 00 78977 A (AUGUST PAUL ;MACNEIL IAN (US); GROSSMAN TRUDY (US); AVENTIS PHARMA)  28 December 2000 (2000-12-28)  page 5, line 1 - line 11  page 11, line 17 - line 19  page 14, line 13 - line 17  claim 21</p> <p style="text-align: center;">--- -/--</p>	<p>1-5,7,8, 18,19</p>



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

7 January 2004

Date of mailing of the international search report

23/01/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Voigt, H

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/07765

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MEJIA J E ET AL: "RETROFITTING VECTORS FOR ESCHERICHIA COLI-BASED ARTIFICIAL CHROMOSOMES (PACS AND BACS) WITH MARKERS FOR TRANSFECTION STUDIES"</p> <p>GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 7, no. 2, February 1997 (1997-02), pages 179-186, XP001106025 ISSN: 1088-9051 page 179, column 2, line 14 - line 21 page 181, column 2, paragraph 2 page 183, column 2, paragraph 1 page 184, column 2, line 20 - line 27 ---</p>	1,2,4,5
X	<p>KIM SU YOUNG ET AL: "Modification of bacterial artificial chromosome clones using Cre recombinase: Introduction of selectable markers for expression in eukaryotic cells."</p> <p>GENOME RESEARCH, vol. 8, no. 4, April 1998 (1998-04), pages 404-412, XP002232358 ISSN: 1088-9051 page 409, column 1, line 3 - line 8 ---</p>	1,4-6
X	<p>DATABASE EMBL 'Online! 23 November 2000 (2000-11-23) "Arabidopsis thaliana chromosom 1" retrieved from EBI Database accession no. AC084820 XP002232359 abstract ---</p>	22
Y	<p>EP 0 176 321 A (LILLY CO ELI) 2 April 1986 (1986-04-02) page 3, line 23 - line 28 ---</p>	1-21
Y	<p>WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30) page 7, line 28 - line 33 page 9, line 8 - line 11 ---</p>	1-21
A	<p>WO 01 40497 A (AVENTIS PHARMA SA ;PERNODET JEAN LUC (FR); COURTOIS SOPHIE (FR); J) 7 June 2001 (2001-06-07) the whole document ---</p>	1-21

-/--



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 03/07765

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>RONDON M R ET AL: "CLONING THE SOIL METAGENOME: A STRATEGY FOR ACCESSING THE GENETIC AND FUNCTIONAL DIVERSITY OF UNCULTURED MICROORGANISMS"</p> <p>APPLIED AND ENVIRONMENTAL MICROBIOLOGY, WASHINGTON, DC, US, vol. 66, no. 6, June 2000 (2000-06), pages 2541-2547, XP000946829</p> <p>ISSN: 0099-2240</p> <p>the whole document</p>	1-21

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-21

Method of analysing a library of polynucleotides comprising the selection of cloning vectors, modification of these vectors and analysis the polynucleotides contained in vectors upon transfer in host cell.

2. Claims: 22-23

Polynucleotide sequence comprising all or part of SEQ ID NO:1 or 2. An oligonucleotide comprising SEQ ID NO:3 or 4.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/EP 03/07765

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 03/07765

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0078977	A	28-12-2000	AU 5496600 A	09-01-2001
			BR 0011765 A	05-03-2002
			CA 2375821 A1	28-12-2000
			EP 1196613 A1	17-04-2002
			JP 2003503029 T	28-01-2003
			NO 20016153 A	15-02-2002
			WO 0078977 A1	28-12-2000
			ZA 200109441 A	17-02-2003
EP 0176321	A	02-04-1986	AU 592956 B2	01-02-1990
			AU 4787785 A	10-04-1986
			CA 1238591 A1	28-06-1988
			DE 3572510 D1	28-09-1989
			DK 429685 A	28-03-1986
			EP 0176321 A1	02-04-1986
			HU 38676 A2	30-06-1986
			IE 58265 B1	25-08-1993
			JP 8015432 B	21-02-1996
			JP 61085191 A	30-04-1986
			US 4921801 A	01-05-1990
WO 9817811	A	30-04-1998	US 5783431 A	21-07-1998
			AU 744960 B2	07-03-2002
			AU 5163298 A	15-05-1998
			CA 2269123 A1	30-04-1998
			EP 0951557 A1	27-10-1999
			JP 2001502547 T	27-02-2001
			KR 2000052795 A	25-08-2000
			WO 9817811 A1	30-04-1998
			US 6242211 B1	05-06-2001
WO 0140497	A	07-06-2001	ZA 9709557 A	12-05-1998
			FR 2801609 A1	01-06-2001
			AU 2179101 A	12-06-2001
			BR 0015993 A	06-08-2002
			CA 2393041 A1	07-06-2001
			EP 1268764 A2	02-01-2003
			WO 0140497 A2	07-06-2001
			JP 2003520578 T	08-07-2003
			NO 20022532 A	26-06-2002

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**